Iowa State University
Digital Repository

Retrospective Theses and Dissertations

Iowa State University Capstones, Theses and Dissertations

1989

Heteroskedasticity-robust estimation of means

Nuwan Nanayakkara
Iowa State University

Follow this and additional works at: https://lib.dr.iastate.edu/rtd

Part of the Statistics and Probability Commons

Recommended Citation

Nanayakkara, Nuwan, "Heteroskedasticity-robust estimation of means" (1989). *Retrospective Theses and Dissertations*. 9230.
https://lib.dr.iastate.edu/rtd/9230

This Dissertation is brought to you for free and open access by the Iowa State University Capstones, Theses and Dissertations at Iowa State University Digital Repository. It has been accepted for inclusion in Retrospective Theses and Dissertations by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.

INFORMATION TO USERS

The most advanced technology has been used to photograph and reproduce this manuscript from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book. These are also available as one exposure on a standard 35mm slide or as a 17" x 23" black and white photographic print for an additional charge.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

U·M·I

University Microfilms International
A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
313/761-4700 800/521-0600

Order Number 8920175

Heteroskedasticity—robust estimation of means

Nanayakkara, Nuwan, Ph.D.

Iowa State University, 1989

U·M·I
300 N. Zeeb Rd.
Ann Arbor, MI 48106

Heteroskedasticity—robust estimation of means

by

Nuwan Nanayakkara

A Dissertation Submitted to the

Graduate Faculty in Partial Fulfillment of the

Requirements for the Degree of

DOCTOR OF PHILOSOPHY

Major: Statistics

Approved:

Signature was redacted for privacy.

In Charge of Major Work

Signature was redacted for privacy.

For the Major Department

Signature was redacted for privacy.

For the Graduate College

Iowa State University
Ames, Iowa

1989

TABLE OF CONTENTS

1. INTRODUCTION AND LITERATURE REVIEW

In univariate statistical analyses we are often confronted with a batch of observations $Y_1, Y_2, \ldots, Y_n$ which are assumed to have an unknown common mean μ, and we wish to perform a test or to construct confidence intervals for this common mean. To be explicit, let us assume that $Y_1, Y_2, \ldots, Y_n$ are observations such that

$$\frac{(Y_i - \mu)}{\sigma_i} \sim G \quad (i = 1, 2, \ldots, n), \tag{1.1}$$

where G has mean 0 and variance 1.

In practice we usually conduct tests or construct confidence intervals for μ using the statistic

$$T = \frac{(\bar{Y} - \mu)}{\sqrt{S^2/n}}, \tag{1.2}$$

where

$$\bar{Y} = \frac{\sum_{i=1}^{n} Y_i}{n},$$

and

$$S^2 = \frac{\sum_{i=1}^{n} (Y_i - \bar{Y})^2}{(n-1)}.$$

Furthermore If we assume that

A(i) Y_i's are independent,

A(ii) σ_i's are equal,

and

A(iii) $G \equiv \Phi$,

where Φ is the standard normal cumulative distribution function, then a remarkable result in statistical theory, established by Student (1908), is that

$$T \sim t_{(n-1)},$$

where $t_{(n-1)}$ is the Student t-distribution with $(n-1)$ degrees of freedom, with density function,

$$f_T(t) = \frac{\Gamma[n/2]}{\sqrt{\pi(n-1)/2}\ \Gamma[(n-1)/2]} \frac{1}{(1+t^2/(n-1))^{n/2}}, \quad -\infty < t < \infty.$$

Now suppose we wish to test the hypotheses

$$H_0: \mu = \mu_0 \quad \text{vs} \quad H_1: \mu \neq \mu_0.$$

Then an α-level test would be to reject H_0 if

$$|T| > t_{n-1,\alpha/2}, \tag{1.3}$$

where $t_{n-1,\alpha/2}$ is the upper $\alpha/2$ percentage point of the t-distribution with $(n-1)$ degrees of freedom. The percentage points are readily available in almost any set of statistical tables. Under the above assumptions A(i), A(ii), and A(iii), one can construct the following $(1-\alpha)100\%$ confidence interval for μ:

$$\left[\bar{Y} - t_{n-1,\alpha/2}\frac{S}{\sqrt{n}},\ \bar{Y} + t_{n-1,\alpha/2}\frac{S}{\sqrt{n}}\right]. \tag{1.4}$$

The validity of the testing procedure given by (1.3) and the confidence interval given by (1.4) certainly depends on the underlying assumptions given by

A(i), A(ii), and A(iii). We should certainly question whether these assumptions are met and investigate the effects of the violations of these assumptions on the distributional properties of T. Assuming $n \geq 2$, $Y_1, Y_2, \ldots, Y_n$ are independent and identically distributed (i.e., A(i) and A(ii) are satisfied), and the common distribution function G has finite moments of all order, Bondesson (1983) shows that T given by (1.2) is t—distributed with $(n-1)$ degrees of freedom if and only if $G = \Phi$. Thus one problem of interest is to determine the effects of the underlying distribution G on the T—statistic; i.e., how robust is T when G is not normal?

In Chapter 2 we shall give results on the effects of violating assumption A(iii). The next paragraphs will be devoted to reviewing important developments in the considerable literature on this subject.

Empirical studies of Neyman and Pearson (1928), "Sophister" (1928) and Nair (1941) show that if G is long (short) tailed compared to the normal distribution, then T tends to be a short (long) tailed. They also show that if G is positively (negatively) skewed, then T tends to be negatively (positively) skewed, and skewness of G affects the distribution of T more than the kurtosis. Hotelling (1961) obtains an expression for the ratio of the tail area of the distribution of T, computed for samples from a known but arbitrary distribution, to the tail area of the usual t—distribution, thus confirming the above findings. From these results we can conclude that, when the symmetry of G is preserved and G is long—tailed compared to the normal distribution, the usual t—test given by (1.3) will be conservative and less powerful, and the confidence interval given by (1.4) will be conservative. Conservativeness of T is also studied by Gross

(1976) and Tukey and McLaughlin (1963). Benjamini (1983) shows the conservativeness of T for long—tailed parent distributions using geometrical arguments similar to the geometrical approaches taken by Hotelling (1961) and Efron (1969). Yuen and Murthy (1974) consider the specific problem of observations drawn from a t—distribution; they tabulate the t values needed for the construction of the confidence limits.

Another approach when sampling from long—tailed distributions is to use Winsorizing or trimming procedures to obtain an estimator for the common mean μ. Andrews et al. (1972), Tukey (1964) and Yuen (1974) consider such procedures. Such estimators can be used to form studentized T—statistics for inferences concerning μ. Tukey and McLaughlin (1963), Huber (1970), and Patel, Mudholkar, and Fernando (1988) consider t—approximations of such studentized T—statistics.

Blachman and Machol (1987) develop confidence limits of the form $\bar{Y} \pm tS/\sqrt{n}$ for the more general location problem with any distribution of known form having unknown location and dispersion, giving particular attention to Cauchy and uniform distributions. In particular they pay attention to the specific t—values to use and tabulate the values of t needed in the construction of the confidence limits for the location parameter if considered as median (mode or mean if it exists) for the specific populations Cauchy, Normal and uniform. For the location problem, Abbott and Rosenblatt (1963) show that there must exist a finite confidence interval for the mean of a normal distribution with unknown variance, based on a single observation with at least any pre—specified probability $0 < 1 - \alpha < 1$. Machol and Rosenblatt (1966) construct the actual confidence interval, for $\alpha < \frac{1}{2}$; they also give confidence limits for the variance of a normal distribution based on a single observation.

Another important departure from assumption A(iii) is the asymmetry of the underlying distribution G. We noted earlier, asymmetry of G affects the distribution of T more than long—tailedness of G. In Chapter 2, we study the effects of asymmetry on the distribution of T using Edgeworth expansions. Bartlett (1935), Geary (1936), Chung (1946), and Gayen (1949) determine the distribution of T by means of Edgeworth or Gram—Charlier expansions. Related results in this area can be found in Bhattacharya and Ghosh (1978) and Callaert and Veraverbeke (1981). Hall (1987) gives an Edgeworth expansion of the T—statistic defined by (1.2), under minimal moment conditions. We shall use the expansion given by Hall (1987) in Section 2.3 to study the effects on the distributional properties of T when assumption A(iii) is violated.

When the distribution of G is skewed, the distribution of T also tends to be skewed, specially in small samples; thus one should not try to approximate the distribution of T by a t—distribution. Since the skewness in the underlying population considerably affects the distribution of T, Johnson (1978) makes a modification for skewness to the T—statistic using Cornish—Fisher expansions. We shall discuss this procedure in detail in Section 2.2. We should also note that these effects decrease as the sample size increase, as a consequence of the central limit theorem. Johnson (1978) and Cressie (1980a) give excellent reviews of the T—statistic, as regards its behavior in the presence of skewness. Cressie, Sheffield, and Whitford (1984) give special attention to the paired comparison t—test on medical data; they give tables for the sample size required to attain a significance level in a specified range, for different levels of skewness and kurtosis of the underlying distribution.

Robustness is a desirable property possessed by some statistical procedures; in words, a robust procedure's performance does not deteriorate badly under departures from a basic set of assumptions. One of the most frequent assumptions that data analysts use is the homogeneity of variances; i.e., that observations are taken with equal precision. The breakdown of this assumption (i.e., violation of assumption A(ii)) is often referred to as heteroskedasticity. To quote Brown (1982), "Indeed, it is fair to say that the topic of robustness of statistical tests against unequal variances is the single most important topic for practical statistics."

When the observations are taken with different precision, intuition tells us to consider a *weighted* average as an estimate for the common mean μ. If we choose nonrandom weights, we should use weights proportional to the inverse of the individual variances, in order to obtain the best (in minimum variance sense) estimate of μ. But this is usually impossible, as the individual variances are not known in practice. Kantorovich (1948) gives an upper bound for the inefficiency of such a weighted average. An accessible proof of this result can be found in Cressie (1980b). If the practitioner wishes to use a pre—assigned set of weights to obtain a weighted average as an estimator of the common mean μ, Cressie (1982) shows how to form a test statistic to test hypotheses concerning the common mean μ. In this paper he addresses the question of misweighting, and uses the notion of "safeness" which we shall discuss in Section 3.4. He also approximates the distribution of this safe T—statistic under the normality assumption by a t—distribution with equivalent degrees of freedom. We shall show that his findings appear again in the simple linear regression problem without intercept, in Section 4.5.

As the optimal set of weights leading to the best estimator of μ is unknown to the practitioner, in the presence of heteroskedasticity it is natural to try random weights to obtain a good estimator of μ. This is possible when the data can be divided into p different identifiable strata such that equal variation occurs within each stratum. If we assume that we have at least 2 observations from each stratum, then one can use the individual sample variances within each stratum to construct a weighted estimator of μ. If we assume that the data are normally distributed then we can easily construct an unbiased estimator of μ with estimated weights. Individual sample means corresponding to the observations coming from each stratum can also be used as unbiased estimators of μ.

Now we should ask "What guarantee is there that the weighted estimator with weights proportional to the inverse of sample variances will be better (in minimum variance sense) than individual sample means?" This question has been addressed by many authors (see discussion below). We can present this problem more formally as follows.

Let $Y_1, Y_2, \ldots, Y_p$ be independent observations such that

$$\frac{Y_i - \mu}{\sigma_i} \sim \Phi \qquad (i = 1, 2, \ldots, p), \tag{1.5}$$

where Φ is the standard normal cumulative distribution function. Let S_i^2 be an estimator of σ_i^2 independent of Y_i, where $\frac{m_i S_i^2}{\sigma_i^2}$ is distributed as chi-square with m_i degrees of freedom ($i = 1,2,\ldots,p$). Now consider the weighted unbiased estimator of μ given by

$$\hat{\mu} = \Big(\sum_{i=1}^{p} \frac{1}{S_i^2} Y_i \Big) \Big/ \Big(\sum_{i=1}^{p} \frac{1}{S_i^2} \Big). \tag{1.6}$$

Then the question is when should we prefer the unbiased estimator $\hat{\mu}$ of μ? That is, when is $\hat{\mu}$ better (usually in minimum variance sense) than the individual Y_i's which are also unbiased? A less general problem of estimating the common mean of two normal populations and the related problem of recovery of interblock information was initiated by Yates (1939, 1940), and his work was extended by Nair (1944) and Rao (1947, 1956). Related work in this area is due to Seshadri (1963,a,b,), Shah (1964), Stein (1966) and Khatri and Shah (1974). Seshadri (1963,b) develops a method of combining interblock and intrablock estimators into an estimator which is uniformly better in the variance sense than either single estimator alone.

Graybill and Deal (1959) prove that for the special case of $p = 2$ strata, $\hat{\mu}$ given by (1.6) is uniformly a better estimator of μ than the individual Y_i's in minimum variance sense if and only if m_1 and m_2 are both greater than 9. Cochran and Carroll (1953) show that when all m_i are equal (say m) and $m > 8$ then as the number of strata $p \to \infty$

$$\hat{\mu} \xrightarrow{d} N\Big[\mu, \frac{(m-2)}{(m-4)} W\Big],$$

where

$$W = \lim_{p \to \infty} \sum_{i=1}^{p} \frac{1}{\sigma_i^2},$$

and for unequal m_i the limiting variance of $\hat{\mu}$ is given by;

$$\lim_{p\to\infty} \frac{\sum_{i=1}^{p} \frac{m_i^2}{(m_i-2)(m_i-4)} \frac{1}{\sigma_i^2}}{\left(\sum_{i=1}^{p} \frac{m_i}{(m_i-2)} \frac{1}{\sigma_i^2}\right)^2}.$$

Meier (1953) gives an approximation to var($\hat{\mu}$) and also gives an unbiased estimator of var($\hat{\mu}$), valid for any p, but neglecting the terms of order $1/m_i^2$. Nair (1980) derives the variance of $\hat{\mu}$ for the special case of $p=2$ strata, as an infinite series. Voinov (1984) generalizes this to any p and presents the exact formulation for var($\hat{\mu}$) using Gauss hypergeometric functions and constructs an unbiased estimator of var($\hat{\mu}$). Voinov further shows that Meier's approximation substantially underestimates the variance of the weighted mean $\hat{\mu}$ when m_i are small or the number of combined groups p is large. This was also observed by Cochran and Carroll (1953) in a sampling investigation where they found that Meier's approximation works well for m_i greater than 10.

Another immediate generalization of the estimator given by (1.6) is to consider an estimator of the type

$$\hat{\mu}^{(2)} = \frac{\sum_{i=1}^{p} \frac{\alpha_i}{S_i^2} Y_i}{\sum_{i=1}^{p} \frac{\alpha_i}{S_i^2}}, \qquad (1.7)$$

where $\alpha_i \geq 0$. We use the notation $\hat{\mu}^{(2)}$ in (1.7) to be consistent with the notation we use in Section 3.3. Clearly, since we assume the independence of Y_i

and S_i^2, unbiasedness of $\hat{\mu}^{(2)}$ immediately follows. This general type of estimators has been studied by many authors, inter alia Norwood and Hinkelmann (1977), Shinozaki (1978), and Bhattacharya (1980, 1984). All these authors give necessary and sufficient conditions for $\hat{\mu}^{(2)}$ to be a uniformly better (in a variance sense) estimator of μ. Kubokawa (1987) considers estimators of type (1.7) and gives sufficient conditions for $\hat{\mu}^{(2)}$ to be a uniformly better estimator under a nondecreasing concave symmetrical loss function.

In Section 3.3 we further generalize (1.7) to consider estimators of the following type.

$$\hat{\mu}^{(r)} = \frac{\sum_{i=1}^{p} \frac{\alpha_i}{S_i^r} Y_i}{\sum_{i=1}^{p} \frac{\alpha_i}{S_i^r}}, \tag{1.8}$$

where $\alpha_i \geq 0$ and $r > 0$. We consider the special case of $p = 2$ and give sufficient conditions (see Theorem 3.3.3) for $\hat{\mu}^{(r)}$ to be uniformly a better estimtor of μ under squared error loss (i.e., in a variance sense). Also we give an upper bound for the inefficiency (see Theorem 3.3.4 (ii)) of $\hat{\mu}^{(2)}$ for this special case of $p = 2$, using a Kantorovich inequality. Two open problems arise here. One is to generalize the sufficient conditions to $p > 2$ and the other is the generalization of the inefficiency upper bound for $p > 2$ and possibly for general r.

Another classical problem in statistical inference is the comparison of two means. This problem is often referred to as the *two-sample problem*. In this situation the experimenter has two batches of observations with common means μ_1 and μ_2. It is of interest to conduct statistical tests or to construct confidence intervals for the difference of means $(\mu_1 - \mu_2)$. Let us assume that

$$\frac{Y_{ij} - \mu_i}{\sigma_{ij}} \sim G_i \qquad (i=1,2,\ j=1,2,\ldots,n_i), \tag{1.9}$$

where G_1 and G_2 have mean 0 and variance 1.

Let

$$\bar{Y}_i = \frac{\sum_{j=1}^{n_i} Y_{ij}}{n_i} \qquad (i=1,2),$$

and

$$S_i^2 = \frac{\sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2}{(n_i - 1)} \qquad (i=1,2).$$

Usual statistical inference for the difference of means $(\mu_1 - \mu_2)$ is performed by considering the statistic T_2^* defined by

$$T_2^* = \frac{\bar{Y}_1 - \bar{Y}_2}{\sqrt{\frac{\{(n_1-1)S_1^2 + (n_2-1)S_2^2\}}{(n_1+n_2-2)} \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}. \tag{1.10}$$

Furthermore if we assume that

B(i) Y_{ij}'s are independent,

B(ii) σ_{ij}'s are equal,

and

B(iii) $G_1 \equiv G_2 \equiv \Phi$,

where Φ is the standard normal cumulative distribution function, then it is immediate from Student's (1908) result that

$$T_2^* \sim t_{(n_1+n_2-2)}. \tag{1.11}$$

Now suppose we wish to test the hypotheses

$$H_0: \mu_1 - \mu_2 = 0 \qquad \text{vs} \qquad H_1: \mu_1 - \mu_2 \neq 0.$$

Then an α-level test would be to reject H_0 if

$$|T_2^*| > t_{n_1+n_2-2,\alpha/2}, \tag{1.12}$$

where $t_{n_1+n_2-2,\alpha/2}$ is the upper $\alpha/2$ percentage point of the t-distribution with (n_1+n_2-2) degrees of freedom. Under the assumptions B(i), B(ii), and B(iii), one can construct the following $(1-\alpha)100\%$ confidence interval for $\mu_1-\mu_2$:

$$(\bar{Y}_1 - \bar{Y}_2) \pm t_{n_1+n_2-2,\alpha/2} \sqrt{\frac{\{(n_1-1)S_1^2+(n_2-1)S_2^2\}}{(n_1+n_2-2)} \left(\frac{1}{n_1}+\frac{1}{n_2}\right)}\ . \tag{1.13}$$

When assumptions B(i), B(ii), and B(iii) no longer hold, what are the distributional properties of T_2^* defined by (1.10)? In what follows we will be mostly concerned with the violations of the latter two assumptions since in most situations (not including time-series data) it is quite reasonable to assume the independence of observations.

Boneau (1960) conducts a series of Monte Carlo studies to see the effects of violations of B(ii) and B(iii). In his work he considers a situation where $\sigma_{ij} = \sigma_i$ (say) $(i=1,2,\ j=1,2,\ldots,n_i)$ but σ_1 may differ from σ_2; i.e., he assumes homogeneity of variances within groups but variances may differ from one group to the other. He concludes that even if the two distributions are not of the same shape but if they are symmetric, then T_2^* is quite robust for such departures from normality. Further he concludes that the differences in skewnesses of the two distributions G_1 and G_2 make the distribution of T_2^* skewed, and in this case inference based on T_2^* will not attain correct significance levels. If the distributions are of the same shape, and sample sizes are unequal but the variances are not too markedly different, then T_2^* is quite robust while combinations of unequal sample sizes and differing variances produce inaccurate probability statements regarding the usual two—sample t—test.

Along the same lines, Havlicek and Peterson (1974) also investigate the effects of violations of assumptions B(ii) and B(iii); i.e., the effects of heterogeneity and nonnormality on the distribution of T_2^*. Using simulation they study these effects separately and in combination and present specific guidelines and tables to assist the experimenter to assess the severity of such violations.

Carter, Khatri, and Srivastava (1979) consider the usual two—sample t—test based on T_2^* (under assumption B(iii)), and conclude that if $\frac{\max\{\sigma_1^2,\sigma_2^2\}}{\min\{\sigma_1^2,\sigma_2^2\}} \leq 1.4$ there is no appreciable effect on the specified significance levels, confirming the findings of Boneau (1960) that we already discussed. They also finds that if one obtains a large number of observations from the population with larger variance then the effects of differing variances seem to be neutralized.

Brown (1982) investigates the effects of unequal variances on a variety of tests used for testing difference of means. He considers the usual two–sample t–tests, and other nonparametric distribution free tests such as sign tests and permutation tests and concludes that robustness of these tests to unequal variances is greatly influenced by unequal sample sizes. He also finds that with equal sample sizes, t–tests are quite robust to unequal variances but the sign test is the most robust test to unequal variances.

The problem of testing for equality of means when $\sigma_{ij}=\sigma_i$ $(i=1,2,$ $j=1,2,\ldots,n_i)$ but $\sigma_1 \neq \sigma_2$ and $G_1=G_2=\Phi$, is called the *Behrens–Fisher problem*. This problem does not have a universally accepted solution. A good account of various types of solutions proposed for this problem are well–documented by Scheffe (1970) and Aucamp (1986). The most commonly used known solution to this problem is given by Welch (1937). Welch considers the statistic T_2 defined by

$$T_2 = \frac{\bar{Y}_1 - \bar{Y}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}, \tag{1.14}$$

where $\bar{Y}_1, \bar{Y}_2, S_1^2$, and S_2^2 are as defined earlier. Here one should note that unless $n_1 = n_2$ or $\sigma_1^2 = \sigma_2^2$, T_2^* does not converge to a *standard* normal distribution whereas T_2 converges to a standard normal distribution regardless of these requirements. This tells us that without some knowledge about how different σ_1^2 is from σ_2^2, one should not try to use T_2^* for inferential problems concerning $\mu_1 - \mu_2$. Welch (1937) showed that T_2 can be approximated by a t–distribution with "equivalent" degrees of freedom (e) given by

$$e = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\left(\frac{S_1^4}{n_1^2(n_1-1)} + \frac{S_2^4}{n_2^2(n_2-1)}\right)}. \tag{1.15}$$

Wang (1971) concludes that one can use T_2 distributed as t_e under H_0 (equality of means) to conduct tests, without much loss of accuracy of the probability of type I error. If one is to use one of the statistics given by T_2^* or T_2, usually a preliminary F—test for the equality of variances σ_1^2 and σ_2^2 is conducted. Gans (1981) compares the tests based on a preliminary F—test and then choice of T_2^* or T_2 depending on the outcome, to the test based on T_2; he concludes that the preliminary F—test for equality of variances is not of much help. Cressie and Whitford (1986) present some rules-of-thumb for when to use the statistic T_2^* as a solution to the Behrens—Fisher problem. Another solution to the Behrens—Fisher problem using the same statistic given by (1.14) is suggested by Cochran (see Cochran and Cox, 1957, p. 101, and Cochran, 1964) and its power characteristics are studied by Lauer (1971).

Aucamp (1986) suggests a new test for the Behrens—Fisher problem and shows that this new test significantly outperforms the usual Z—test based on T_2^* when sample sizes n_1 and n_2 are large. In the light of our discussion just above, about the lack of robustness of T_2^*, this should not be surprising.

From Boneau's (1960) investigations we see that the differences in skewness greatly influence the distributions of T_2^* and T_2; see also Cressie and Whitford (1986). Thus, following Johnson's (1978) approach of correcting for skewness using Cornish—Fisher expansions in the one—sample problem, Cressie and

Whitford (1986) also do the same for the two-sample problem. They also obtain a formula to assess the effects of differing population skewnesses on T_2^* and T_2 and use Posten's (1979) tables to assess these effects.

The testing of equality of several means of different populations from independent samples is another common statistical problem which falls under the framework of one-way analysis of variance (ANOVA). This problem includes the two-sample problem as a special case.

Before we proceed let us consider the usual one-way fixed effects model of analysis of variance. One can write this model more formally as

$$Y_{ij} = \mu_i + \epsilon_{ij} \qquad (i=1,2,\ldots,p,\ j=1,2,\ldots,n_i), \tag{1.16}$$

where it is assumed that $\frac{\epsilon_{ij}-0}{\sigma_{ij}} \sim G_i$ $(i=1,2,\ldots,p,\ J=1,2,\ldots,n_i)$, and G_i has mean zero and variance 1. Furthermore the following assumptions are usually made.

C(i) ϵ_{ij}'s are independent,

C(ii) σ_{ij}'s are equal,

and

C(iii) $G_i \equiv \Phi \quad (i=1,2,\ldots,p)$,

where Φ is the standard normal cumulative distribution function.

Just as in the one-sample and two-sample problems, the equality of variances assumption (i.e., C(ii)) has a greater effect on the analysis than small deviations from the normality assumption (i.e., assumption C(iii)).

It is well known (see, Scheffe, 1959) that tests based on the one—way ANOVA F—statistic are sensitive to lack of homogeneity of within group variances. That is, the actual size of a test is greatly infuenced by different underlying population variances. Box (1954), Box and Anderson (1955) and Box and Watson (1962) consider robustness of these analyses to unequal variances, under normal or nearly normal errors; they conclude that if the design is balanced, i.e., the n_i's are equal, the usual tests are quite robust to unequal variances as long as the sample sizes are not too small. They also conclude that the unbalanced designs do not acquire this property. In other words, specified significance levels are affected by unequal variances in unbalanced designs. This is referred to as the Box principle (e.g., see Brown, 1982).

Many authors, inter alia Welch (1951), James (1951), Banerjee (1960), Ury and Wiggins (1971), Spjotvoll (1972), Brown and Forsythe (1974), Games and Howell (1976), Hochberg (1976), Tamhane (1977,1979), and Dalal (1978) have focused their attention on the multiple sample problem in the presence of unequal variances under normality. Banerjee (1960) develops a confidence interval for any linear function of the form $\sum_{i=1}^{p} C_i \mu_i$ (where C_i are known constants) with confidence coefficient not less than the pre—assigned probability of coverage. Brown and Forsythe (1974) develop a new F-type statistic which is similar to the usual ANOVA F—statistic except for a small denominator correction that takes unequal variances into account; they use Satterthwaite's (1946) approximation to obtain the denominator degrees of freedom. They compare the performance of this new statistic to the usual F—statistic, a statistic proposed by Welch (1951) and a statistic proposed by James (1951), via a Monte Carlo sampling experiment. Their results show that the usual F—statistic is greatly influenced by strong heterogeneity among variances, and that the other three are quite robust to such situations. They also conclude that when the population variances are equal, or nearly equal their critical region of the suggested F—type statistic more closely approximates that of the usual ANOVA F than does Welch's statistic.

Tamhane (1977) proposes single—stage procedures for (i) all pairwise comparisons and all linear contrasts among the means μ_i and (ii) all linear combinations of the means μ_i. These procedures are based on Banerjee's (1961) method and Welch's (1937) method. He conducts a Monte Carlo simulation to study these two procedures and shows that both procedures guarantee the specified probability of coverage of .90 or .95 but the procedure based on Welch's method fails to guarantee the specified probability of 0.99 in some cases. These simulations also show that the procedure based on Banerjee's method is highly conservative.

Surprisingly, the effects of the violation of assumption C(iii) on the one—way ANOVA has not been studied until very recently. Tan and Tabatabai (1986) conduct a simulation study to see the effects of unequal variances in combinations with nonnormality on the test suggested by Welch (1951), James (1951) and Brown and Forsythe (1974). Their results show that each of the three tests above are quite robust to departures from normality and the differences among these tests are so small that the choice is immaterial for practical purposes.

One—sample, two—sample, and one—way ANOVA problems can be put under a broader framework which is known as the linear model. The analyses based on linear model theory is valid under assumptions that are given below.

Let

$$Y_i = \underline{X}_i'\underline{\beta} + \epsilon_i \qquad (i=1,2,\ldots,n), \tag{1.17}$$

where $\{\epsilon_i : i=1,2,\ldots,n\}$ is a sequence of independently distributed random errors such that $\frac{\epsilon_i - 0}{\sigma_i} \sim G_i$ with G_i having mean 0 and variance 1, $\underline{\beta}$ is an unknown vector of parameters of length k; and $\underline{X}_i'$ is a k x 1 vector of deterministic components (fixed regressors).

Our interest lies primarily in estimating the unknown parameter vector $\underline{\beta}$ and making inference on $\underline{\beta}$. In standard linear model theory the following assumptions are usually made.

D(i) ϵ_i's are independent,

D(ii) σ_i's are equal,

and

D(iii) $G_i \equiv \Phi \qquad (i=1,2,\ldots,n)$,

where Φ is the standard normal cumulative distribution function. The assumptions D(ii) and D(iii) are often referred to as the homoskedasticity and the normality assumptions respectively.

As linear model theory plays an important role in statistics and more generally in our everyday life through its applications in social sciences, physics, engineering, geology, etc., we should ask what the consequences are of violating these assumptions. Since nature is not as smooth as our model, practioners who handle real—life data are always encountering situations where these assumptions are violated. What is one to do if violations occur?

Searching for an answer to this question has led many authors to investigate the consequences of departures from the assumptions above and to suggest inferential procedures robust to such departures.

It is not uncommon to find violation of the homoskedasticity assumption. For example, Prais and Houthakker (1955) find in their study of family budgets that the variability of expenditures has an increasing trend as household income increase. Other data sets where the homoskedasticity assumption is violated can also be found in Hinkley (1977), Carroll and Ruppert (1982), Rutemiller and Bowers (1968) and Koenker and Bassett (1982).

Henceforth we shall refer to the model (1.17) with assumption D(ii) violated as a heteroskedastic linear model. Such models are commonly used in fields including economics, biological sciences, and physical sciences.

It is well known that ordinary least squares theory under heteroskedastic models leads to consistent but often inefficient parameter estimates and inconsistent covariance matrix estimates and that these effects are not minor; see Geary (1966), and Goldfeld and Quandt (1972), Chapter 3. If we knew the structure of the heteroskedasticity we might overcome the difficulty by performing a suitable transformation of the data. But this knowledge is often not at hand. As we discuss in the one—sample problem of estimating a common mean μ (Chapter 3), it is sensible to perform a weighted least squares analysis if we believe that the homoskedasticity assumption is violated. When the different σ_i's are known to the practitioner then he or she can proceed with a weighted least squares analysis which is optimal under the normality assumption. In practice of course, the σ_i^2's are usually unknown.

The two most common methods of handling heteroskedastic models is to assume

(i) replication at design points.

or

(ii) variance is a continuous function of known form depending on $\underline{X}_i$, $\underline{\beta}$, and some unknown parameters.

Examples of authors who have used assumption (i), are: Bement and Williams (1969), Williams (1967), Fuller and Rao (1978), Deaton, Reynolds and Myers (1983) and Carroll and Cline (1988); examples of authors who have used assumption (ii), are: Rutemiller and Bowers (1968), Amemiya (1973), Box and Hill (1974), Bickel (1978), Jobson and Fuller (1980), Carroll and Ruppert (1982), Cook and Weisberg (1982), Davidian and Carroll (1987), and Anh (1988).

Bement and Williams (1969) assume normality of errors and use sample variances as weights to perform a weighted regression analysis. They apply this method of estimated weighted least squares (e.w.l.s.) to four common problems: two— and multiple—sample problems, and simple linear regression with and without intercept; they conclude that the number of replicates at each design point must be at least 10 in order to obtain good results from e.w.l.s. compared to the unweighted least squares. They also provide an asymptotically correct formula for the variance of the e.w.l.s. estimator. The same suggestion about the number of replicates was also made by Williams (1975). Jacquez, Mather, and Crawford (1968) and Jacquez and Norusis (1973) conduct simulation studies empirically verifying this suggestion. A simulation study of Deaton, Reynolds, and Myers (1983) especially conducted for the simple linear heteroskedastic regression model shows that the above minimum number of 10 replicates at each design point depends upon the severity of variance heterogeneity and they give more specific guidelines for when to use e.w.l.s.

Rao (1970) proposes an estimator for the unknown covariance matrix of the error terms. This is usually known as a MINQU estimator. One can use this estimator to perform a weighted least squares regression instead of weighting by the usual sample variances. Rao and Subrahmanian (1971), Jacquez and Norusis (1973), Rao (1973) and Chaubey and Rao (1976) study the relative merits of MINQUE based estimates, sample variance based estimates, and ordinary least squares estimates and conclude that for many cases of interest MINQUE based estimates outperform the other two.

Carroll and Cline (1988) use two weighting schemes: the usual sample variances as Bement and Williams (1969) do, and sample average squared residuals from a preliminary regression fit. They show that for asymmetrically distributed data, the weighted least squares estimates are generally inconsistent and if the number of replicates equals 2 at each design point, then even under normality the e.w.l.s. estimates based on sample variances are inconsistent. Asymptotic normality of both estimates is proved and the superiority of the weights obtained from a preliminary regression fit over weights inversly proportional to the usual sample variances, is demonstrated for the special case of normally distributed data.

Carroll and Ruppert (1982) assume $\sigma_i = H(\underline{X}_i, \underline{\beta}, \underline{\theta})$ where H is a smooth known function and $\underline{\theta}$ is an unknown parameter vector, and they show the existence of a wide class of robust estimators of $\underline{\beta}$. They prove that as long as a reasonable estimator of $\underline{\theta}$ is available their estimators of $\underline{\beta}$ are asymptotically equivalent to the natural estimates obtained via weighted least squares with known σ_i's. In addition, they propose a method of obtaining a reasonable estimate of $\underline{\theta}$. Anh (1988) also assumes the smoothness of the variance function as above. In particular he assumes that $\sigma_i = \sigma |\underline{X}_i' \underline{\beta}|^{\gamma}$, where σ and γ are unknown, and arrives at a set of nonlinear equations using $\hat{\underline{\beta}}$, the o.l.s. estimate, as an initial estimate of $\underline{\beta}$ and proceeds to obtain estimates of $\sigma, \underline{\beta}$, and γ. These estimates are then used as initial estimates in an iterative maximum likelihood scheme to derive more efficient estimates of the unknown parameters.

In the simple linear regression model when the variances are proportional to a power of the mean, Miller (1986) suggests the use of empirical weights to obtain weighted least squares estimates of the unknown slope and intercept parameters; i.e., use weights estimated by the inverse of the appropriate power of the response variable, something which is quick and easy to do in practice. Dorfman (1988) conducts a simulation study to investigate the effects of this procedure on the bias of regression estimates and on the coverage probabilities of the associated confidence intervals. He concludes that inference on the slope parameter is reasonably good when the variance is proportional to the mean; as the proportionality constant grows, confidence levels deteriorate and point estimates of both the parameters, slope and intercept tend to be negatively biased.

One might consider another heteroskedastic situation where equal variance occurs except at a few random design points where the variablity may be very large. Many robust—regression techniques have been proposed to guard against such gross errors; see Carroll (1980), Belsey, Kuh, and Welsch (1980), Huber

(1981), Bickel and Doksum (1981), and Birch and Binkley (1983).

Dalal, Tukey, and Cohen (1984) combine these robust regression techniques with an assumption that the error variances are locally smooth functions of the regressor variables except for a few points that are suspected as outliers; they use simple linear regression. Since the robust techniques we mentioned earlier guard against the undue influences of outliers, their procedure smooths nonoutlying residuals from a robust regression fit and hence obtain weights for a weighted regression. They show using a Monte Carlo simulation, that their technique is better than the usual robust regression methods in the presence of heterogeneity, but that it does not perform well when the variances are equal.

We discussed various techniques proposed by many authors which can be used in specific situations, but is there a more general approach? As we noted earlier, heteroskedasticity leads to inefficient ordinary least squares estimates rendering their estimated standard errors inconsistent. Thus under heterogeneity of variances one cannot use ordinary least squares theory to make valid inference even asymptotically. Alternate approaches for consistently estimating the covariance matrix of the ordinary least squares estimator of $\underline{\beta}$ even under heteroskedasticity have been suggested by Eicker (1963), Hartley, Rao and Kiefer (1969), Chew (1970), Rao (1970), Hinkley (1977), White (1980a), and MacKinnon and White (1985). The estimators proposed by Chew and Rao are not only consistent but also unbiased. Other estimators are generally biased although asymptotically unbiased. In Chapter 4 we shall discuss the applications of White's results to one—sample, two—sample, and simple linear regression problems in connection with Rao's, and MacKinnon and White's estimators.

A more general approach in modelling is to consider nonlinear regression models. Formally we can write the model as follows:

$$Y_i = f(\underline{X}_i, \underline{\beta}) + \epsilon_i \qquad (i=1,2,\ldots,n), \tag{1.18}$$

where f is a known function of regressor variables $\underline{X}_i$ and unknown parameter vector $\underline{\beta}$. Our interest is to make inferences about the unknown parameter vector $\underline{\beta}$.

Jenrich (1969), Malinvaud ((1970), Wu (1979) considered models of type (1.18) with fixed (nonstochastic) regressors $\underline{X}_i$ and independent and identically distributed errors and give sufficient conditions for the consistency and normality of the nonlinear least squares estimator of the unknown parameter vector $\underline{\beta}$. Shao (1988) considers fixed regressors and independent but not necessarily identically distributed errors and gives sufficient conditions for the consistency and asymptotic normality of the nonlinear least squares estimator. Hannan (1971) extends Jenrich's (1969) results to time—series data. White (1980b) extends Jenrich's results to the case of stochastic regressors and assumes that errors are independent of regressors but not necessarily identically distributed. White and Domowitz (1984) consider a similar model with heteroskedastic and/or serially correlated errors. They also give general conditions to ensure consistency and asymptotic normality for the nonlinear least squares estimator and propose a new covariance matrix estimator that is consistent regardless of the heteroskedasticity or serial correlation of unknown form . In this dissertation we shall only be considering linear models.

2. ONE SAMPLE T—STATISTIC

2.1. Introduction

In Chapter 1 we discussed the basic problems covered in this dissertation. Briefly, in this chapter we are interested in making inference about the population mean μ when the underlying distribution deviates from normality. Based on often overly optimistic assumptions about how the data were generated, one typically makes inference about the population mean μ by assuming that the T—statistic has a Student's t—distribution. Bondesson (1983) shows that the usual T—statistic is t—distributed if and only if the underlying population is normally distributed. So naturally we should ask "Are there any normal populations? What if the population is not normal? Can we still use the usual t—test safely or perhaps with some modifications?" Cressie et al. (1984) examine the consequences of departures from normality in the t—test and conclude that with some qualifications the t—test is quite immune to such departures.

Johnson (1978) and Cressie (1980a) give excellent reviews of the 1—sample T—statistic. Early empirical studies of Neyman and Pearson (1928), "Sophister" (1928), and Nair (1941) show that positive (negative) skewness in the population results in negative (positive) skewness in the distribution of the usual T—statistic. Also their studies show that skewness of the underlying population affects the distribution of T more than the kurtosis, and long—tailed parent populations causes T to be shorter—tailed than for the normal parent.

Thus when the underlying distribution is long—tailed compared to the normal distribution, the usual Student's t—test is conservative and less powerful. Benjamini (1983) establishes this fact using geometrical arguments.

In Section 2.2 we shall discuss a modification for skewness of the T—statistic using Cornish—Fisher expansions. This modification was suggested by Johnson (1978); we correct the misprints in Appendix A of Johnson's article. In Section 2.3 we give a new approach to study the usual T—statistic, using Edgeworth expansions.

Before we proceed, the necessary notations will be introduced. Let $Y_1, Y_2, \ldots, Y_n$ be a random sample from a population with mean μ and let $\sigma^2, \mu_3, \mu_4, \ldots$ represent the second, third, fourth, $\cdots$ central moments of the underlying population.

Define

$$Sk = \frac{\mu_3}{\sigma^3}, \tag{2.1}$$

$$Ku = \frac{\mu_4}{\sigma^4}, \tag{2.2}$$

$$\bar{Y} = \frac{\sum_{i=1}^{n} Y_i}{n}, \tag{2.3}$$

$$S^2 = \frac{\sum_{i=1}^{n} (Y_i - \bar{Y})^2}{(n-1)}, \tag{2.4}$$

and

$$T = \frac{(\bar{Y} - \mu)}{\sqrt{\frac{S^2}{n}}}. \tag{2.5}$$

2.2. Modification of the T—statistic (Johnson 1978)

For any random variable Y the general form of the Cornish—Fisher expansion is given by (Cornish and Fisher, 1937)

$$CF(Y) = \mu + \sigma\varsigma + \frac{\mu_3}{6\sigma^2}(\varsigma^2 - 1) + \cdots, \tag{2.6}$$

where ς is a standard normal random variable. Wallace (1958) discusses the validity of such series approximations of a random variable. Discussions on Cornish—Fisher expansions can also be found in Ord (1972, pp. 32—34), Kendall and Stuart (1963, pp. 165—166).

Now using (2.6) we obtain

$$CF(\bar{Y}) = \mu + \frac{\sigma}{\sqrt{n}}\varsigma + \frac{\mu_3}{6n\sigma^2}(\varsigma^2 - 1) + A_n, \tag{2.7}$$

where $A_n = O_p(n^{-3/2})$; i.e., for every $\epsilon > 0$ there exist a constant $K(\epsilon)$ and an integer $n(\epsilon)$ such that, if $n \geq n(\epsilon)$ then $\Pr\left\{\frac{|A_n|}{n^{-3/2}} \leq K(\epsilon)\right\} \geq 1 - \epsilon$.

We notice here that in the above expansion, the skewness of the population, μ_3, is the coefficient of the $(\varsigma^2 - 1)$ term. In fact it also appears in the coefficients of other terms, but they are of smaller order. The key in obtaining a modified T—variable in Johnson's approach is to eliminate the term involving μ_3 in the general T—variable defined below.

Let the general T—variable be

$$T_J = \frac{(\bar{Y} - \mu) + \lambda + \gamma\{(\bar{Y} - \mu)^2 - (\sigma^2/n)\}}{\sqrt{S^2/n}}. \tag{2.8}$$

The numerator of (2.8) is suggested by looking at the inverse Cornish—Fisher expansion of ς in terms of $(\bar{Y}-\mu)$ in (2.6); λ in T_J is chosen so that the constant terms in the Cornish—Fisher expansion of T_J sum to zero so that the lower—order bias is eliminated, and γ is chosen so that the coefficient of (ς^2-1) term in the Cornish—Fisher expansion of T_J is zero (thereby eliminating the lower—order effects of skewness). We give the derivation of λ and γ below.

It can be shown easily that

$$E(S^2) = \sigma^2, \tag{2.9}$$

$$\mathrm{var}(S^2) = \frac{\mu_4}{n} - \frac{(n-3)}{n(n-1)}\sigma^4$$

$$= \frac{(\mu_4-\sigma^4)}{n} + \frac{2\sigma^4}{n(n-1)}.$$

Hence we obtain

$$(\mathrm{var}(S^2))^{1/2} = \left(\frac{\mu_4-\sigma^4}{n}\right)^{1/2}\left[1+\frac{2\sigma^4}{(n-1)(\mu_4-\sigma^4)}\right]^{1/2}$$

$$= \left(\frac{\mu_4-\sigma^4}{n}\right)^{1/2}\left[1+\frac{2\sigma^4}{(n-1)(\mu_4-\sigma^4)}+O(n^{-2})\right]$$

$$= \left(\frac{\mu_4-\sigma^4}{n}\right)^{1/2} + O(n^{-1}). \tag{2.10}$$

Now using (2.6), (2.9), and (2.10) we obtain the Cornish—Fisher expansion of S^2, viz.

$$CF(S^2) = \sigma^2 + \left(\frac{\mu_4-\sigma^4}{n}\right)^{1/2}\eta + O_p(n^{-1}), \tag{2.11}$$

so that

$$CF(S^2/n)^{-1/2} = \frac{n^{1/2}}{\sigma}\left\{1-\frac{1}{2}\left(\frac{\mu_4-\sigma^4}{n\sigma^4}\right)^{1/2}\eta\right\} + O_p(n^{-1}), \tag{2.12}$$

where η is a standard normal random variable.

Substituting (2.7) and (2.12) in (2.8) we obtain

$$\mathrm{CF}(T_J) = \Big[\frac{\sigma}{n^{1/2}}\varsigma + \frac{\mu_3}{6n^{1/2}\sigma^2}(\varsigma^2-1) + \lambda + \gamma\frac{\sigma^2}{n}(\varsigma^2-1)\Big]\cdot$$

$$\Big\{\frac{n^{1/2}}{\sigma}\Big[1 - \frac{1}{2}\Big(\frac{\mu_4-\sigma^4}{n\sigma^4}\Big)^{1/2}\eta\Big] + O_p(n^{-1})\Big\}$$

$$= \varsigma + \frac{\mu_3}{6n^{1/2}\sigma^2}(\varsigma^2-1) + \frac{\lambda n^{1/2}}{\sigma} + \frac{\gamma\sigma}{n^{1/2}}(\varsigma^2-1)$$

$$-\frac{1}{2}\Big(\frac{\mu_4-\sigma^4}{n\sigma^4}\Big)^{1/2}\eta\varsigma + O_p(n^{-1}). \qquad (2.13)$$

One should notice here that ς and η are standard normal random variables, but they are correlated; ς appears in the Cornish–Fisher expansion of $\bar{Y}$ and η appears in the Cornish–Fisher expansion of S^2.

Let

$$\rho = \mathrm{corr}(\bar{Y}, S^2).$$

It can be shown after some algebra that

$$\rho = \frac{\mu_3}{n}\Big(\frac{\sigma^2}{n}\cdot\frac{(n-1)\mu_4-(n-3)\sigma^4}{n(n-1)}\Big)^{-1/2}$$

$$= \mu_3\Big[\sigma^2(\mu_4-\sigma^4) + \frac{2\sigma^6}{(n-1)}\Big]^{-1/2}$$

$$= \mu_3\Big[\sigma^2(\mu_4-\sigma^4)\Big]^{-1/2}\Big[1 + \frac{2\sigma^4}{(\mu_4-\sigma^4)(n-1)}\Big]^{-1/2}$$

$$= \mu_3\Big[\sigma^2(\mu_4-\sigma^4)\Big]^{-1/2} + O(n^{-1}). \qquad (2.14)$$

Now let

$$\eta = \rho\varsigma + (1-\rho^2)^{1/2}\varsigma^*, \tag{2.15}$$

where ς^* is a standard normal random variable independent of ς.
Substituting (2.14) and (2.15) in (2.13) we obtain

$$\begin{aligned}
CF(T_J) &= \varsigma + \frac{\mu_3}{6n^{1/2}\sigma^3}(\varsigma^2-1) + \frac{\lambda n^{1/2}}{\sigma} + \frac{\gamma\sigma}{n^{1/2}}(\varsigma^2-1) \\
&\quad -\frac{1}{2}\left(\frac{\mu_4-\sigma^4}{n\sigma^4}\right)^{1/2}\frac{\mu_3}{\left(\sigma^2(\mu_4-\sigma^4)\right)^{1/2}}\varsigma^2 \\
&\quad -\frac{1}{2}\left(\frac{\mu_4-\sigma^4}{n\sigma^4}\right)^{1/2}\left(\frac{\sigma^2(\mu_4-\sigma^4)-\mu_3^2}{\sigma^2(\mu_4-\sigma^4)}\right)^{1/2}\varsigma\varsigma^* + O_p(n^{-1}) \\
&= \varsigma + \left(\frac{\mu_3}{6n^{1/2}\sigma^3} + \frac{\gamma\sigma}{n^{1/2}} - \frac{\mu_3}{2n^{1/2}\sigma^3}\right)(\varsigma^2-1) + \frac{\lambda n^{1/2}}{\sigma} \\
&\quad -\frac{\mu_3}{2n^{1/2}\sigma^3} - \frac{1}{2n^{1/2}\sigma^3}\left(\sigma^2(\mu_4-\sigma^4)-\mu_3^2\right)^{1/2}\varsigma\varsigma^* + O_p(n^{-1}).
\end{aligned}$$

Now setting the coefficient of the (ς^2-1) term equal to zero and setting the sum of the constant terms equal to zero we obtain

$$\gamma = \frac{\mu_3}{3\sigma^4},$$

and

$$\lambda = \frac{\mu_3}{2n\sigma^2}.$$

The γ and λ given above yield

$$T_J = \frac{(\bar{Y}-\mu)+\mu_3/6\sigma^2 n+(\mu_3/3\sigma^4)(\bar{Y}-\mu)^2}{(S^2/n)^{1/2}}, \tag{2.16}$$

and

$$CF(T_J) = \varsigma - \frac{1}{2n^{1/2}}(Ku-1-Sk^2)^{1/2}\varsigma\varsigma^* + O_p(n^{-1}), \tag{2.17}$$

where Sk and Ku are as defined in (2.1) and (2.2).

We see that T_J given by (2.16) is not computable since μ_3 and σ^2 are usually unknown. Johnson suggests replacing μ_3 and σ^2 by the usual sample estimates: $\hat{\mu}_3=\{\sum_{i=1}^{n}(Y_i-\bar{Y})^3\}/n$ and sample variance S^2 respectively; the Cornish–Fisher expansion is still (2.17). To demonstrate the use of the statistic assume we wish to test the hypotheses

$$H_0:\mu = \mu_0 \quad \text{vs} \quad H_1:\mu > \mu_0.$$

An α–level test would be to reject H_0 if

$$T_{JC} = \frac{(\bar{Y}-\mu)+\hat{\mu}_3/6S^2 n+(\hat{\mu}_3/3S^4)(\bar{Y}-\mu)^2}{(S^2/n)^{1/2}} > t_{\alpha,\nu},$$

where the value $t_{\alpha,\nu}$ is obtained from the Student's t–distribution with $\nu=(n-1)$ degrees of freedom and α is the probability that the t–random variable is greater than $t_{\alpha,\nu}$.

Suppose now we ask the question, "After the adjustment is made to the T–variable is it still reasonable to use $(n-1)$ degrees of freedom for its approximate t—distribution?" We could answer this question as follows. Suppose T_J follows an approximate t—distribution with f degrees of freedom. Then we should be able to write the Cornish—Fisher expansion of T_J as

$$CF(T_J) = Z - \frac{1}{(2f)^{1/2}} ZZ^* + O_p(f^{-1}), \tag{2.18}$$

where Z and Z^* are independent standard normal random variables.
Now comparing (2.17) and (2.18) we obtain

$$\frac{1}{(2f)^{1/2}} = \frac{1}{2n^{1/2}}(Ku - 1 - Sk^2)^{1/2},$$

and hence

$$f = \frac{2n}{(Ku - 1 - Sk^2)}.$$

The same result holds true for T_{JC}. One should note here a well known result that for any population, the quantity $(Ku - 1 - Sk^2)$ is nonnegative; see for example Kendall and Stuart (1963, p. 92). Again for practical purposes Ku and Sk can be estimated by the sample observations and hence we obtain

$$\hat{f} = \frac{2n}{(\hat{K}u - 1 - \hat{S}k^2)}, \tag{2.19}$$

where

$$\hat{S}k = \frac{\{\sum_{i=1}^{n}(Y_i - \bar{Y})^3\}/n}{\{\sum_{i=1}^{n}(Y_i - \bar{Y})^2/n\}^{3/2}},$$

and

$$\hat{K}u = \frac{\{\sum_{i=1}^{n}(Y_i - \bar{Y})^4\}/n}{\{\sum_{i=1}^{n}(Y_i - \bar{Y})^2/n\}^2}.$$

In the next section, we shall examine the properties of the usual T—statistic using an Edgeworth (1898) expansion of the T—statistic and draw some conclusions. Reviews about Edgeworth expansions of Student's T—statistic are well documented by Wallace (1958), Bowman, Beauchamp and Shenton (1977) and Cressie (1980a). Hall (1987) gives an expansion of the T—statistic under minimal moment conditions. We will be using Hall's expansion in the next section. At this point we should also note that the expansions obtained by Bhattacharya and Ghosh (1978), when applied to the T—statistic give an expansion with a remainder $o(n^{-k/2})$, provided we assume that the underlying population has finite $2(k+2)^{th}$ order moment. Hall's work allows us to obtain an expansion with a remainder $o(n^{-k/2})$ assuming only the finiteness of the $(k+2)^{th}$ moment and nonsingularity of the underlying distribution.

2.3. Edgeworth Expansion of the T—statistic and Related Results

In this section, we will use the following theorem proved by Hall (1987) in order to draw some conclusions and to give some recommendations on the use of the T—statistic.

2.3.1. Theorem (Hall, 1987)

Assume $k \geq 1$, $E|Y|^{k+2} < \infty$ and the distribution of Y is nonsingular. Then

$$F_{T_o}(y) = \Pr(T_o \leq y) = \Phi(y) + \sum_{i=1}^{n} n^{-i/2} P_i(y)\phi(y) + o(n^{-k/2}), \tag{2.20}$$

uniformly in y. A function F is called singular if and only if it is not identically zero and F' (exists and) equals zero a.e. Here the P_i's are the polynomials of degree $(3i-1)$ appearing in the formal Edgeworth expansion of the distribution of T_O; see below. Φ and ϕ are the standard normal cumulative and density functions respectively, and

$$T_O = n^{-1/2}(\bar{Y}-\mu)(n^{-1}\sum_{i=1}^{n} Y_i^2 - \bar{Y}^2)^{-1/2}.$$

□

The coefficients P_i are functions of $\mu, \sigma^2, \mu_3, \ldots, \mu_{i+2}$. For example:

$$P_1(y) = \tfrac{1}{6}Sk(2y^2+1),$$

and

$$P_2(y) = -y\{\tfrac{1}{18}Sk^2(y^4+2y^2-3)-\tfrac{1}{12}(Ku-3)(y^2-3)+\tfrac{1}{4}(y^2+3)\},$$

where Sk and Ku are the measures of skewness and kurtosis of the population defined as in (2.1) and (2.2) respectively.

We immediately notice that

$$T = \left(\frac{n-1}{n}\right)^{1/2} T_O, \tag{2.21}$$

where T is the usual T-statistic as defined in (2.5). From (2.20) for the special case of $k=2$ we obtain

$$\Pr(T_O \le 0) = \tfrac{1}{2} + \frac{n^{-1/2}}{6(2\pi)^{1/2}}Sk + o(n^{-1}). \tag{2.22}$$

In what follows in this section we will assume the finiteness of the fourth moment and nonsingularity of the underlying distribution. Now we will derive the first few moments of T_O and hence obtain the first few moments of T.

$$E(T_O) = \frac{n^{-1/2}}{6} Sk \int_{-\infty}^{\infty} y\, d[(2y^2+1)\phi(y)]$$

$$-n^{-1} \int_{-\infty}^{\infty} y\, d[(\tfrac{1}{18} Sk^2 y(y^2+3)(y^2-1))\phi(y)]$$

$$+n^{-1} \int_{-\infty}^{\infty} y\, d[\{\tfrac{1}{12}(Ku-3)y(y^2-3)+\tfrac{1}{4}y(y^2+3)\}\phi(y)] + o(n^{-1}).$$

After performing integration, it is a simple matter to show that

$$E(T_O) = -\frac{n^{-1/2}}{2} Sk + o(n^{-1}). \tag{2.23}$$

We see immediately from (2.21)—(2.23) that as the sample size n increases, no matter what the value of the population skewness Sk is, T_O and hence T becomes more and more symmetric about 0 (i.e., as $n \to \infty$, the median of $T \to 0$ and the mean of $T \to 0$); see Groeneveld and Meeden (1977) for a discussion on distribution symmetry. This should not be surprising since it is well known that $T \xrightarrow{d} Z$, where Z is a standard normal random variable. Thus for moderately large sample sizes or only slightly skewed populations, T_O and hence T is approximately symmetric about 0.

Also we can show after some simple integration

$$E(T_O^2) = 1 + \tfrac{2}{n} Sk^2 + \tfrac{3}{n} + o(n^{-1}), \tag{2.24}$$

and hence

$$\mathrm{var}(T_O) = 1 + \frac{7}{4n}Sk^2 + \frac{3}{n} + o(n^{-1}). \tag{2.25}$$

Now from (2.21) we obtain similar relationships for T.

$$\Pr(T \leq 0) = \frac{1}{2} + \frac{n^{-1/2}}{6(2\pi)^{1/2}}Sk + o(n^{-1}), \tag{2.26}$$

$$E(T) = -\frac{n^{-1/2}}{2}Sk + o(n^{-1}), \tag{2.27}$$

and

$$\mathrm{var}(T) = 1 + \frac{7}{4n}Sk^2 + \frac{2}{n} + o(n^{-1}). \tag{2.28}$$

Now consider the Taylor series expansion of $F_{T_O}(y)$ around $y = 0$, obtained from (2.20). It can be shown that

$$F_{T_O}(y) = \frac{1}{2} + \frac{n^{-1/2}}{6(2\pi)^{1/2}}Sk + \frac{y}{(2\pi)^{1/2}}\left(1 + \frac{Sk^2}{6n} - \frac{Ku}{4n}\right) + \frac{n^{-1/2}}{4(2\pi)^{1/2}}y^2\,Sk + o(n^{-1}), \tag{2.29}$$

and hence using (2.21) we obtain

$$F_T(y) = \frac{1}{2} + \frac{n^{-1/2}}{6(2\pi)^{1/2}}Sk + \frac{y}{(2\pi)^{1/2}}\left(1 - \frac{1}{2n} + \frac{Sk^2}{6n} - \frac{Ku}{4n}\right) + \frac{n^{-1/2}}{4(2\pi)^{1/2}}y^2\,Sk + o(n^{-1}). \tag{2.30}$$

To find the median of the distribution of T we have to solve $F_T(y) = \frac{1}{2}$. An approximate feasible solution to this equation is $y = \frac{n^{-1/2}}{6} Sk$ up to $O(n^{-1})$. Therefore up to $O(n^{-1})$ we have the following

$$\text{Median of } T = \frac{n^{-1/2}}{6} Sk, \tag{2.31}$$

and

$$\text{Mean of } T = -\frac{n^{-1/2}}{2} Sk. \tag{2.32}$$

From (2.31) and (2.32) we see that, if the parent distribution is skewed to the right (i.e., $Sk > 0$) then $-\frac{n^{-1/2}}{2} Sk < 0 < \frac{n^{-1/2}}{6} Sk$; thus the distribution of T is skewed to the left. The converse is true for $Sk < 0$. These conclusions agree with the previously mentioned studies cited in the introduction of this chapter and Chapter 1. Thus we can conclude that, when the parent distribution is heavily skewed and sample size is small the skewness of the distribution of T means we should not try to approximate it by a t—distribution unless we make a correction for skewness such as done by Johnson (1978).

Henceforth in this section we shall concentrate on symmetric parent distributions, i.e., $Sk = 0$. When we sample from a normal parent distribution (for which $Sk = 0$ and $Ku = 3$) we know that T follows a t—distribution with $(n-1)$ degrees of freedom. Also we concluded earlier that long—tailed parent distributions give rise to short—tailed distributions for T and vice versa. Hence our intuition tell us that if we sample from a long (short) tailed parent distribution compared to the normal distribution, and if we try to approximate the distribution of T by a t—distribution with some appropriate degrees of freedom, then the equivalent degrees of freedom are likely to be greater (smaller) than $(n-1)$.

To show this for symmetric distributions, it is necessary to match fourth moments.

$$E(T_o^4) = 3 - \tfrac{2}{n}(Ku - 3) + \tfrac{24}{n} + o(n^{-1}),$$

and therefore

$$E(T^4) = 3 - \tfrac{2}{n}(Ku - 3) + \tfrac{18}{n} + o(n^{-1}).$$

In the case of a normal parent distribution, since we know that

$$E(T^4) = \frac{3(n-1)^2}{(n-3)(n-5)},$$

we could write in general for any symmetric distribution

$$E(T^4) \approx \frac{3(n-1)^2}{(n-3)(n-5)} - \tfrac{2}{n}(Ku - 3).$$

Now, let T follow an approximate t–distribution with f degrees of freedom. Then the equivalent degrees of freedom f can be obtained by solving,

$$\frac{3(f-1)^2}{(f-3)(f-5)} = \frac{3(n-1)^2}{(n-3)(n-5)} - \tfrac{2}{n}(Ku - 3). \tag{2.33}$$

We can easily see that

$$f > (n-1) \qquad \text{if } Ku > 3,$$

and

$$f < (n-1) \qquad \text{if } Ku < 3.$$

The above result shows that if we sample from a long—tailed distribution then the correspoding T—statistic is short—tailed and vice versa. Also we showed earlier that positive skewness in the parent population causes T to be negatively skewed and vice versa. These observations certainly agree with the previously cited literature in Chapter 1 and the introduction of this chapter.

3. WEIGHTED ESTIMATION OF A LOCATION PARAMETER

3.1. Introduction

In Chapter 2, using Cornish—Fisher expansions we discussed some modifications of the usual T—statistic for skewness of the population. Also, using Edgeworth expansions we gave some recommendations on the use of the T—statistic when observations are drawn from any population with finite third and fourth moments. In this chapter, we will relax the homoskedasticity assumption made in the previous chapter and proceed towards greater generality. Specifically, assume that the independent random variables (i.e., sample observations) $Y_1, Y_2, \ldots, Y_n$ are such that $(Y_i - \mu)/\sigma_i \sim G$; $i = 1, 2, \ldots, n$, where the cumulative distribution function (c.d.f.) G could be standard normal c.d.f Φ or will have mean 0 and variance 1.

In most situations, we are interested in estimating (point estimation or interval estimation) the unknown common mean μ or in testing hypotheses about the common mean μ. We would like a point estimate of μ to be unbiased and to have small variance. First, we should understand that each observation Y_i contains some information about the unknown mean μ but with different precision due to the differences in the variances. This naturally suggests the use of a weighted sample mean $\bar{Y}_W = \sum_{i=1}^{n} w_i Y_i$, where $\sum_{i=1}^{n} w_i = 1$, to estimate the common mean μ. Then what are the optimal weights $\{w_i\}$ to use? Section 3.2 will address this question. Section 3.3 will be devoted to the presentation of some results on weighted estimation of the common mean μ, while Section 3.4 will contain a method of forming a "safe" T—statistic T_W for arbitrary weights in $\bar{Y}_W$, introduced by Cressie (1982). Extension of these ideas to M—estimation will be developed in Section 3.5.

3.2. Optimal Weights

For the rest of this chapter, let $Y_1, Y_2, \ldots, Y_n$ be such that $\{(Y_i - \mu)/\sigma_i\}$ are identically distributed with mean 0 and variance 1.

Define

$$\bar{Y}_W = \sum_{i=1}^{n} w_i Y_i, \tag{3.1}$$

where

$$\sum_{i=1}^{n} w_i = 1.$$

Clearly, $\sum_{i=1}^{n} \bar{Y}_W$ is an unbiased estimator of the common mean μ, and

$$\mathrm{var}(\bar{Y}_W) = \sum_{i=1}^{n} w_i^2 \sigma_i^2. \tag{3.2}$$

We would like to choose the weights $\{w_i,\ i=1,2,\ldots,n\}$ such that $\mathrm{var}(\bar{Y}_W)$ is minimized subject to the constraint $\sum_{i=1}^{n} w_i = 1$. It can be shown easily using the method of Langrange multipliers that $\mathrm{var}(\bar{Y}_W)$ is minimized when $w_i \propto 1/\sigma_i^2$. Moreover, this choice of weights maximizes the asymptotic power in testing problems concerning the common mean μ. This can be seen as follows. Suppose we are interested in testing

$$H_0: \mu \le \mu_0 \qquad \text{vs} \qquad H_1: \mu > \mu_0.$$

Assume the σ_i's are known and let

$$Z_W = \frac{\bar{Y}_W - \mu}{(\sum_{i=1}^{n} w_i^2 \sigma_i^2)^{1/2}}.$$

Furthermore, assume that

$$Z_W \xrightarrow{d} Z,$$

where typically Z is a standard normal random variable and $\xrightarrow{d}$ indicates convergence in distribution. (The exact distribution of Z_W is usually unknown, but if it is known, then the arguments given below remain valid for finite sample situations.) Then asymptotically an α $-$level test would be to reject H_o if

$$Z_W = \frac{\bar{Y}_W - \mu}{(\sum_{i=1}^{n} w_i^2 \sigma_i^2)^{1/2}} > z_\alpha,$$

where z_α is such that $\Pr(Z > z_\alpha) = \alpha$. Let $\mu_1 = \mu_o + \Delta$. Then

$$\begin{aligned} \text{Power}(\mu_1) &= \Pr(\text{reject } H_o\text{, when } \mu = \mu_1) \\ &= \Pr\left\{\frac{\bar{Y}_W - \mu_o}{(\sum_{i=1}^{n} w_i^2 \sigma_i^2)^{1/2}} > z_\alpha\text{, when } \mu = \mu_1\right\} \\ &= \Pr\left\{\frac{\bar{Y}_W - (\mu_o + \Delta) + \Delta}{(\sum_{i=1}^{n} w_i^2 \sigma_i^2)^{1/2}} > z_\alpha\text{, when } \mu = \mu_1\right\}. \end{aligned}$$

Now assuming that $\sum_{i=1}^{n} w_i^2 \sigma_i^2$ converges to $\sum_{i=1}^{\infty} w_i^2 \sigma_i^2$, as $n \to \infty$, we obtain

$$\text{Power}(\mu_1) \to \Pr\left(Z + \frac{\Delta}{(\sum_{i=1}^{\infty} w_i^2 \sigma_i^2)^{1/2}} > z_\alpha\right), \text{ as } n \to \infty.$$

Therefore, asymptotic power can be maximized by maximizing $\Delta/(\sum_{i=1}^{\infty} w_i^2\sigma_i^2)^{1/2}$; i.e., by minimizing $(\sum_{i=1}^{\infty} w_i^2\sigma_i^2)^{1/2}$ subject to the constraint $\sum_{i=1}^{\infty} w_i = 1$. This is achieved when $w_i \propto 1/\sigma_i^2$.

3.3. Weighted Estimation of a Common Mean μ

In the previous section, we discussed the importance of weighted estimation of a common mean μ when the observations are heteroskedastic. One objective of considering weighted estimators is to obtain an unbiased estimator of μ that is more efficient (i.e., smaller variability). The following theorem due to Kantorovich (1948), gives an upper bound for the inefficiency of a weighted linear unbiased estimator of μ (i.e., the ratio of the variance of a linear unbiased estimator of μ to the optimally weighted estimator of μ).

3.3.1. Theorem (Kantorovich, 1948)

Let $\hat{\mu}_W = \bar{Y}_W = \sum_{i=1}^{n} w_i Y_i$ be a weighted unbiased estimator of μ; where $w_i > 0$, $i = 1,2,\ldots,n$, and $\sum_{i=1}^{n} w_i = 1$. Let $\hat{\mu}_O = \sum_{i=1}^{n} w_i^O Y_i$, where $w_i^O = (1/\sigma_i^2)/(\sum_{i=1}^{n} 1/\sigma_i^2)$, be the optimally weighted unbiased estimator of μ. Then

$$\frac{\mathrm{var}(\hat{\mu}_W)}{\mathrm{var}(\hat{\mu}_O)} \le \frac{(R+1)^2}{4R}, \tag{3.3}$$

where

$$R = \frac{\max\{w_i\sigma_i^2,\ i = 1,2,\ldots,n\}}{\min\{w_i\sigma_i^2,\ i = 1,2,\ldots,n\}}.$$

□

We shall omit the proof of the above theorem and refer the reader to Cressie (1980b) for an accessible proof. He also gives references to multivariate generalizations of the theorem.

In what follows in this section, we shall assume that the data can be divided into p identifiable strata such that equal variances occur within each stratum. In practice, this occurs when say p laboratory technicians make duplicate measurements of a certain characteristic, μ (e.g., length, weight, etc.) of an object, all using the same instrument. The variability in the measurements that one technician will make certainly depend on his or her skills (and the instrument being used) and thus it is a sensible thing to model these p groups of observations as having possibly different variances. Each group of observations will provide us with an (unbiased) estimator of μ. Our aim is to combine these estimators to arrive at a more efficient estimator of μ. Combining two such estimators is a common problem that arises in applied statistics. This problem is particularly important in combining inter—block and intra—block estimators in the incomplete block design, see for example Bhattacharya (1980).

Let

$$\frac{Y_{ij}-\mu}{\sigma_i} \sim G \qquad (i=1,2,\ldots,p \text{ and } j=1,2,\ldots,n_i), \tag{3.4}$$

where G has mean 0 and variance 1,

$$\hat{\mu}_i = \bar{Y}_i = \frac{\sum_{j=1}^{n_i} Y_{ij}}{n_i} \qquad (i=1,2,\ldots,p), \tag{3.5}$$

and

$$\hat{\sigma}_i^2 = S_i^2 = \frac{\sum_{j=1}^{n_i}(Y_{ij} - \bar{Y}_i)^2}{(n_i - 1)} \qquad (j = 1,2,\ldots,p). \tag{3.6}$$

Clearly, $\hat{\mu}_i$ is an unbiased estimator of μ, $\hat{\sigma}_i^2$ is an unbiased estimator of σ_i^2 and $\text{var}(\hat{\mu}_i) = \sigma_i^2/n_i$; $i = 1,2,\ldots,p$. Thus, in the light of the discussion of the previous section, the linear combined estimator of μ that is unbiased and has minimum variance is given by

$$\hat{\mu}_O = \frac{\sum_{i=1}^{p}(n_i/\sigma_i^2)\hat{\mu}_i}{\sum_{i=1}^{p}(n_i/\sigma_i^2)}, \tag{3.7}$$

and

$$\text{var}(\hat{\mu}_O) = \frac{1}{\sum_{i=1}^{p}(n_i/\sigma_i^2)}. \tag{3.8}$$

In general the $\{\sigma_i^2\}$ are unknown so this estimator is not of practical use. If $\{w_i\}$ is any set of constants with the properties $0 \le w_i \le 1$ and $\sum_{i=1}^{p} w_i = 1$, then $\hat{\mu}_W = \sum_{i=1}^{p} w_i \hat{\mu}_i$ is a linear unbiased estimator of μ. It is well known (see Graybill and Deal, 1959) that for any such set of fixed weights, the variance of $\hat{\mu}_W$ is greater than either σ_1^2/n_1 or σ_2^2/n_2 or...or σ_p^2/n_p for some choices of parameters $\sigma_1^2, \sigma_2^2, \ldots, \sigma_p^2$. Thus, there is no set of fixed weights $\{w_i\}$ such that the estimator $\hat{\mu}_W$ has smaller variance than $\hat{\mu}_1$ and $\hat{\mu}_2 \cdots$ and $\hat{\mu}_p$, uniformly over the parameter space of variances.

3.3.2. Definition

Let the weights $\{v_i\}$ be such that $\sum_{i=1}^{p} v_i = 1$, $0 \le v_i \le 1$ and $\hat{\alpha}_V = \sum_{i=1}^{p} v_i \hat{\alpha}_i$ be an unbiased estimator of α where $\hat{\alpha}_i$; $i = 1,2,\ldots,p$, are also unbiased estimators of α with variances β_i^2; $i=1,2,\ldots,p$, respectively. If the variance of $\hat{\alpha}_V$ is less than or equal to $\min\{\beta_i^2;\ i = 1,2,\ldots,p\}$ for all possible values of β_i^2; $i = 1,2,\ldots,p$, then $\hat{\alpha}_V$ is called a uniformly better unbiased estimator of α.

Since there exists no such set of constant weights $\{w_i;\ i = 1,2,\ldots,p\}$ that will give rise to a uniformly better unbiased estimator of μ, one might try using random weights. We shall now prove the following theorem for the special case of two strata (i.e., $p = 2$) and normal distribution (i.e., $G = \Phi$).

3.3.3. Theorem

Let

$$\frac{Y_{ij} - \mu}{\sigma_i} \sim \Phi \qquad (i = 1,2;\ j = 1,2,\ldots,n_i),$$

where Φ is the standard normal cumulative distribution function, and Y_{ij} is independent of Y_{kl} for all possible i,j,k, and l. Let $r > 0$, and n_1, n_2 be integers such that $n_1, n_2 > 2r + 1$, and

$$R = \frac{\max\{\sigma_1^2/n_1, \sigma_2^2/n_2 : \sigma_1^2, \sigma_2^2 \in \Sigma\}}{\min\{\sigma_1^2/n_1, \sigma_2^2/n_2 : \sigma_1^2, \sigma_2^2 \in \Sigma\}}, \quad \Sigma \subseteq \mathbb{R}^+ \setminus \{0, \infty\}.$$

Then, a sufficient condition for

$$\hat{\mu}^{(r)} = \frac{\left[\frac{a_1}{S_1^r}\bar{Y}_1 + \frac{a_2}{S_2^r}\bar{Y}_2\right]}{\left[\frac{a_1}{S_1^r} + \frac{a_2}{S_2^r}\right]}, \tag{3.9}$$

to be a uniformly better unbiased estimator of μ than $\bar{Y}_1$ and $\bar{Y}_2$ is that a_1 and a_2 should be chosen in such a way that

$$\frac{R^{(2-r)/2}}{2}\left[\frac{n_1(n_2-1)}{n_2(n_1-1)}\right]^{r/2}\frac{B\left[\frac{n_1-1+2r}{2},\frac{n_2-1-2r}{2}\right]}{B\left[\frac{n_1-1+r}{2},\frac{n_2-1-r}{2}\right]} \leq \frac{a_1}{a_2}$$

$$\leq \frac{2}{R^{(2-r)/2}}\left[\frac{n_1(n_2-1)}{n_2(n_1-1)}\right]^{r/2}\frac{B\left[\frac{n_1-1-r}{2},\frac{n_2-1+r}{2}\right]}{B\left[\frac{n_1-1-2r}{2},\frac{n_2-1+2r}{2}\right]}, \qquad \text{for } 0<r<2, \quad (3.10)$$

$$\frac{1}{2}\left[\frac{n_1(n_2-1)}{n_2(n_1-1)}\right]^{r/2}\frac{B\left[\frac{n_1-1+2r}{2},\frac{n_2-1-2r}{2}\right]}{B\left[\frac{n_1-1+r}{2},\frac{n_2-1-r}{2}\right]} \leq \frac{a_1}{a_2}$$

$$\leq 2\left[\frac{n_1(n_2-1)}{n_2(n_1-1)}\right]^{r/2}\frac{B\left[\frac{n_1-1-r}{2},\frac{n_2-1+r}{2}\right]}{B\left[\frac{n_1-1-2r}{2},\frac{n_2-1+2r}{2}\right]}, \qquad \text{for } r\geq 2, \quad (3.11)$$

and in particular when $r=2$ the condition given in (3.11) becomes

$$\frac{n_1(n_1+1)(n_2-1)}{2(n_1-1)n_2(n_2-5)} \leq \frac{a_1}{a_2} \leq \frac{2n_1(n_1-5)(n_2-1)}{(n_1-1)n_2(n_2+1)}, \qquad (3.12)$$

where $B(\cdot,\cdot)$ is the beta function given by $B(a,b)=\frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$,

$$\bar{Y}_i = \frac{\sum_{j=1}^{n_i} Y_{ij}}{n_i},$$

and

$$S_i^2 = \frac{\sum_{j=1}^{n_i}(Y_{ij}-\bar{Y}_i)}{(n_i-1)} \qquad (i=1,2).$$

Proof:

First we observe the following

$$E(\bar{Y}_i) = \mu \qquad (i = 1,2), \tag{3.13}$$

and $(\bar{Y}_1, \bar{Y}_2, S_1, S_2)$ are mutually independent. This immediately shows the unbiasedness of $\hat{\mu}^{(r)}$. Now we can write

$$\begin{aligned}
\operatorname{var}(\hat{\mu}^{(r)}) &= E(\hat{\mu}^{(r)} - \mu)^2 \\
&= E\left(\frac{(a_1/S_1^r)\bar{Y}_1 + (a_2/S_2^r)\bar{Y}_2}{(a_1/S_1^r + a_2/S_2^r)} - \mu\right)^2 \\
&= E\left(\frac{a_1/S_1^r}{a_1/S_1^r + a_2/S_2^r}(\bar{Y}_1 - \mu) + \frac{a_2/S_2^r}{a_1/S_1^r + a_2/S_2^r}(\bar{Y}_2 - \mu)\right)^2 \\
&= E\left(\frac{a_1/S_1^r}{a_1/S_1^r + a_2/S_2^r}\right)^2 \frac{\sigma_1^2}{n_1} + E\left(\frac{a_2/S_2^r}{a_1/S_1^r + a_2/S_2^r}\right)^2 \frac{\sigma_2^2}{n_2}.
\end{aligned}$$

If we let

$$\gamma = \frac{a_1/S_1^r}{(a_1/S_1^r + a_2/S_2^r)},$$

then

$$\operatorname{var}(\hat{\mu}^{(r)}) = E(\gamma^2)\frac{\sigma_1^2}{n_1} + E(1-\gamma)^2\frac{\sigma_2^2}{n_2}. \tag{3.14}$$

Observe that

$$\begin{aligned}
\gamma &= 1/\left(1 + \frac{a_2}{a_1}\frac{S_1^r}{S_2^r}\right) \\
&= 1/\left(1 + \frac{a_2}{a_1}\frac{\sigma_1^r}{\sigma_2^r}\left(\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}\right)^{r/2}\right) \\
&= 1/(1 + cF^{r/2}),
\end{aligned} \tag{3.15}$$

where $c = \frac{a_2}{a_1}\frac{\sigma_1^r}{\sigma_2^r}$ and $F \sim F_{n_1-1, n_2-1}$, an F—distribution on $n_1 - 1$ and $n_2 - 1$ degrees of freedom.

Therefore from (3.14) and (3.15)

$$\operatorname{var}(\hat{\mu}^{(r)}) = \frac{\sigma_1^2}{n_1} E\left[\frac{1}{(1+cH)^2}\right] + \frac{\sigma_2^2}{n_2} E\left[\frac{c^2H^2}{(1+cH)^2}\right], \tag{3.16}$$

where $H = F^{r/2}$.

Let

$$k = \frac{\sigma_1^2/n_1}{\sigma_2^2/n_2}.$$

<u>Case I</u>: $\sigma_1^2/n_1 \le \sigma_2^2/n_2$ i.e., $k \in [\frac{1}{R}, 1]$.

Then we would like to choose a_1 and a_2 in such a way that

$$\operatorname{var}(\hat{\mu}^{(r)}) \le \sigma_1^2/n_1 \quad \text{for all } k \in [\tfrac{1}{R}, 1].$$

From (3.16)

$$\operatorname{var}(\hat{\mu}^{(r)}) = \frac{\sigma_1^2}{n_1}\left[E\left(\frac{1}{(1+cH)^2}\right) + \frac{1}{k}E\left(\frac{c^2H^2}{(1+cH)^2}\right)\right].$$

Therefore, we need to choose a_1 and a_2 in such a way that

$$E\left(\frac{1}{(1+cH)^2}\right) + \frac{1}{k}E\left(\frac{c^2H^2}{(1+cH)^2}\right) \le 1, \quad \text{for all } k \in [\tfrac{1}{R}, 1].$$

i.e.,

$$E\left(\frac{k + c^2H^2}{k(1+cH)^2}\right) \le 1, \quad \text{for all } k \in [\tfrac{1}{R}, 1].$$

Now, since $c = \frac{a_2}{a_1}\frac{\sigma_1^r}{\sigma_2^r}$ and $k = \frac{\sigma_1^2/n_1}{\sigma_2^2/n_2}$ we obtain $c = \frac{a_2}{a_1}\left(\frac{n_1}{n_2}k\right)^{r/2}$ and thus we need to

choose a_1 and a_2 in such a way that

$$E\left[\frac{k+\frac{\alpha_2^2}{\alpha_1^2}\left(\frac{n_1}{n_2}k\right)^r H^2}{k\left\{1+\frac{\alpha_2}{\alpha_1}\left(\frac{n_1}{n_2}k\right)^{r/2}H\right\}^2}\right] \leq 1, \quad \text{for all } k \in [\tfrac{1}{R},1].$$

Substitute $N_1 = n_1^{r/2}$, $N_2 = n_2^{r/2}$, $l = k^{r/2}$ and $s = 2(r-1)/r$. So we need to choose α_1 and α_2 in such a way that

$$E\left[\frac{\alpha_1^2 N_2^2 + \alpha_2^2 N_1^2 l^s H^2}{(\alpha_1 N_2 + \alpha_2 N_1 l H)^2}\right] \leq 1, \quad \text{for all } l \in [\frac{1}{R^{r/2}},1]. \tag{3.17}$$

<u>Case</u> <u>II</u>: $\sigma_2^2/n_2 \leq \sigma_1^2/n_1$ i.e., $k \in [1,R]$.

Here we should choose α_1 and α_2 in such a way that

$$\mathrm{var}(\hat{\mu}^{(r)}) \leq \sigma_2^2/n_2, \quad \text{for all } k \in [1,R].$$

Again, from (3.16) and writing $k^* = 1/k$, $l^* = k^{*^{r/2}}$ and $H^* = 1/H$, we see that in this case by a similar consideration we should choose α_1 and α_2 in such a way that

$$E\left[\frac{\alpha_2^2 N_1^2 + \alpha_1^2 N_2^2 l^{*^s} H^{*^2}}{(\alpha_2 N_1 + \alpha_1 N_2 l^* H^*)^2}\right] \leq 1, \quad \text{for all } l^* \in [\frac{1}{R^{r/2}},1]. \tag{3.18}$$

Since we do not know whether σ_1^2/n_1 is larger or smaller than σ_2^2/n_2 we should find α_1 and α_2 such that (3.17) and (3.18) are both satisfied.

Now, consider (3.17). Let

$$h(H) = \frac{\alpha_1^2 N_2^2 + \alpha_2^2 N_1^2 l^s H^2}{(\alpha_1 N_2 + \alpha_2 N_1 l H)^2}.$$

We immediately observe the following:

$$h(0)=1,\ h(1)=\frac{a_1^2N_2^2+a_2^2N_1^2l^S}{(a_1N_2+a_2N_1l)^2}.$$

and

$$h'(H)=\Big[(a_1N_2+a_2N_1lH)^2 2a_2^2N_1^2l^SH-(a_1^2N_2^2+a_2^2N_1^2l^SH^2)\cdot$$

$$2a_2N_1l(a_1N_2+a_2N_1lH)\Big]/(a_1N_2+a_2N_1lH)^4.$$

Therefore, $h'(H)=0$ when $H=\dfrac{a_1N_2l}{a_2N_1l^S}$ and $h\Big(\dfrac{a_1N_2l}{a_2N_1l^S}\Big)=\dfrac{1}{(1+l^{2-S})}$.

Now let us approximate $h(H)$ by the parabola

$$U(H)=\nu+\delta(H-\theta)^2,$$

which we will center at $H=\dfrac{a_1N_2l}{a_2N_1l^S}$, and force to pass through $(0,1)$ and $\Big(\dfrac{a_1N_2l}{a_2N_1l^S},\dfrac{1}{(1+l^{2-S})}\Big)$. Then it is a simple matter to show that

$$U(H)=\frac{1}{(1+l^{2-S})}+\frac{l^Sa_2^2N_1^2}{a_1^2N_2^2(1+l^{2-S})}\Big(H-\frac{a_1N_2l}{a_2N_1l^S}\Big)^2.$$

Now we will show that $h(H)\le U(H)$ for all H; i.e., we want to show that

$$\frac{a_1^2N_2^2+a_2^2N_1^2l^SH^2}{(a_1N_2+a_2N_1lH)^2}\le\frac{1}{(1+l^{2-S})}+\frac{l^Sa_2^2N_1^2}{a_1^2N_2^2(1+l^{2-S})}\Big(H-\frac{a_1N_2l}{a_2N_1l^S}\Big)^2. \tag{3.19}$$

Substitute $a_2N_1H=a_1N_2Z$. Then we can show after some algebra that the above inequality (3.19) is satisfied if

$$\frac{1+l^SZ^2}{(1+lZ)^2}\le 1+\frac{l^S}{(1+l^{2-S})}\Big(Z^2-\frac{2lZ}{l^S}\Big);$$

i.e., if

$$\frac{l^S Z^2 - 2lZ - l^2 Z^2}{(1+lZ)^2} \le \frac{l^S}{(1+l^{2-S})}\Big(Z^2 - \frac{2lZ}{l^S}\Big);$$

i.e., if

$$\frac{l^S}{(1+l^{2-S})}\Big(Z - \frac{2l}{l^S}\Big) - \frac{l^S Z - 2l - l^2 Z}{(1+lZ)^2} \ge 0;$$

i.e., if

$$\frac{(1+lZ)^2 l^S (Z - 2l^{1-S}) - (1+l^{2-S})(l^S Z - 2l - l^2 Z)}{(1+l^{2-S})(1+lZ)^2} \ge 0. \tag{3.20}$$

Now since $(1+l^{2-S})(1+lZ)^2 > 0$, we need to show only the numerator in (3.20) is nonnegative. After some algebra one can show that

$$\begin{aligned}&(1+lZ)^2 l^S (Z - 2l^{1-S}) - (1+l^{2-S})(l^S Z - 2l - l^2 Z)\\&= l^{S+1}(Z + lZ)(Z - l^{1-S})^2 \ge 0,\end{aligned}$$

and thus

$$h(H) \le U(H), \text{ for all } H.$$

Hence

$$E(h(H)) \le E(U(H)).$$

Now

$$E((U(H)) = 1 + \frac{l^S a_2^2 N_1^2}{a_1^2 N_2^2 (1+l^{2-S})}\Big\{E(H^2) - \frac{2a_1 N_2 l}{a_2 N_1 l^S} E(H)\Big\}.$$

From the above equation it is clear that if we choose a_1 and a_2 in such a way that

$$E(H^2) - \frac{2a_1 N_2 l}{a_2 N_1 l^S} E(H) \le 0 \qquad \text{for all } l \in [\frac{1}{R^{r/2}}, 1],$$

then (3.17) will be satisfied. That is, we need to choose a_1 and a_2 in such a way that

$$\frac{a_1}{a_2} \geq \frac{N_1 E(H^2)}{2N_2 l^{1-s} E(H)};$$

i.e.,

$$\frac{a_1}{a_2} \geq \frac{n_1^{r/2}}{2n_2^{r/2} k^{(2-r)/2}} \frac{E(F^r)}{E(F^{r/2})}, \quad \text{for all } k \in [\tfrac{1}{R}, 1].$$

Now we observe the following:

$$\text{if } 0 < r < 2 \quad \text{then} \quad R^{(2-r)/2} \geq \frac{1}{k^{(2-r)/2}}, \quad \text{for all } k \in [\tfrac{1}{R}, 1],$$

and

$$\text{if } r \geq 2 \text{ then} \quad 1 \geq \frac{1}{k^{(2-r)/2}} \quad \text{for all } k \in [\tfrac{1}{R}, 1].$$

So if we choose a_1 and a_2 in such a way that

$$\frac{a_1}{a_2} \geq \frac{n_1^{r/2} R^{(2-r)/2}}{2n_2^{r/2}} \frac{E(F^r)}{E(F^{r/2})} \quad \text{if } 0 < r < 2,$$

and

$$\frac{a_1}{a_2} \geq \frac{n_1^{r/2}}{2n_2^{r/2}} \frac{E(F^r)}{E(F^{r/2})} \quad \text{if } r \geq 2,$$

then (3.17) will be satisfied. Now since $F \sim F_{n_1-1,n_2-1}$, where F_{n_1-1,n_2-1} is Snedecor's F—distribution, we obtain

$$E(F^{r/2}) = \left[\frac{n_2-1}{n_1-1}\right]^{r/2} \frac{B\left[\frac{n_1-1+r}{2}, \frac{n_2-1-r}{2}\right]}{B\left[\frac{n_1-1}{2}, \frac{n_2-1}{2}\right]},$$

and

$$E(F^r) = \left[\frac{n_2-1}{n_1-1}\right]^{r} \frac{B\left[\frac{n_1-1+2r}{2}, \frac{n_2-1-2r}{2}\right]}{B\left[\frac{n_1-1}{2}, \frac{n_2-1}{2}\right]}.$$

Therefore if we choose α_1 and α_2 in such a way that

$$\frac{\alpha_1}{\alpha_2} \geq \frac{R^{(2-r)/2}}{2} \left[\frac{n_1(n_2-1)}{n_2(n_1-1)}\right]^{r/2} \frac{B\left[\frac{n_1-1+2r}{2},\frac{n_2-1-2r}{2}\right]}{B\left[\frac{n_1-1+r}{2},\frac{n_2-1-r}{2}\right]},$$

$$\text{for } 0<r<2, \tag{3.21}$$

and

$$\frac{\alpha_1}{\alpha_2} \geq \frac{1}{2}\left[\frac{n_1(n_2-1)}{n_2(n_1-1)}\right]^{r/2} \frac{B\left[\frac{n_1-1+2r}{2},\frac{n_2-1-2r}{2}\right]}{B\left[\frac{n_1-1+r}{2},\frac{n_2-1-r}{2}\right]} \quad \text{for } r\geq 2, \tag{3.22}$$

then (3.17) will be satisfied. Similarly considering (3.18) one can show that if we choose α_1 and α_2 in such a way that

$$\frac{\alpha_1}{\alpha_2} \geq \frac{1}{2R^{(2-r)/2}}\left[\frac{n_1(n_2-1)}{n_2(n_1-1)}\right]^{r/2} \frac{B\left[\frac{n_1-1+2r}{2},\frac{n_2-1-2r}{2}\right]}{B\left[\frac{n_1-1+r}{2},\frac{n_2-1-r}{2}\right]} \quad \text{for } 0<r<2, \tag{3.23}$$

and

$$\frac{\alpha_1}{\alpha_2} \geq 2\left[\frac{n_1(n_2-1)}{n_2(n_1-1)}\right]^{r/2} \frac{B\left[\frac{n_1-1+2r}{2},\frac{n_2-1-2r}{2}\right]}{B\left[\frac{n_1-1+r}{2},\frac{n_2-1-r}{2}\right]} \quad \text{for } r\geq 2, \tag{3.24}$$

then (3.18) will be satisfied. Now combining (3.21)—(3.24) we obtain the required result.

For the special case $r=2$ we immediately notice that

$$\frac{B\left[\frac{n_1-1+2r}{2},\frac{n_2-1-2r}{2}\right]}{B\left[\frac{n_1-1+r}{2},\frac{n_2-1-r}{2}\right]} = \frac{B\left[\frac{n_1+3}{2},\frac{n_2-5}{2}\right]}{B\left[\frac{n_1+1}{2},\frac{n_2-3}{2}\right]}$$

$$= \frac{\Gamma\left(\frac{n_1+3}{2}\right)\Gamma\left(\frac{n_2-5}{2}\right)}{\Gamma\left(\frac{n_1+1}{2}\right)\Gamma\left(\frac{n_2-3}{2}\right)} = \frac{(n_1+1)}{(n_2-5)},$$

where $\Gamma(\cdot)$ is the gamma function.

Now substituting this in (3.22) and (3.24), we obtain the required result.

Q.E.D.

Many authors have considered only $\hat{\mu}^{(2)}$, a special case of the unbiased estimator given by (3.9) (i.e., $r=2$), and we shall now summarize those contributions. Graybill and Deal (1959) show that if a_1 and a_2 in the definition of $\hat{\mu}^{(2)}$ are chosen to be n_1 and n_2 respectively, then $\hat{\mu}^{(2)}$ will be a uniformly better unbiased estimator of μ iff n_1 and n_2 are both greater than 10. Norwood and Hinkelmann (1977), Shinozaki (1978) and Bhattacharya (1984) consider unbiased estimators of the type given by (3.9) for the special case of $r=2$, not only for two strata but more generally for p strata and give necessary and sufficient conditions for this estimator to be a uniformly better unbiased estimator of μ than the individual sample means. Kubokawa (1987) generalizes these results and gives sufficient conditions for the combined estimator to have a smaller risk than each sample mean with respect to a nondecreasing concave loss function. We should also note here that under squared error loss for the special case of $r=2$, the conditions given by (3.12) also become necessary.

In the next theorem we present the exact expression for $\mathrm{var}(\hat{\mu}^{(2)})$ and also give an upper bound for the inefficiency of the estimator given by $\hat{\mu}^{(2)}$ (i.e., the ratio of $\mathrm{var}(\hat{\mu}^{(2)})$ to the variance of the optimally weighted unbiased estimator) using the Kantorovich inequality.

3.3.4. Theorem

Let the notations be as in Theorem 3.3.3, and $\hat{\mu}^{(2)}$ be as defined by (3.9). Let $\hat{\mu}_O$ be the optimally weighted unbiased estimator of μ. That is to say

$$\hat{\mu}_O = \frac{(n_1/\sigma_1^2)\bar{Y}_1 + (n_2/\sigma_2^2)\bar{Y}_2}{(n_1/\sigma_1^2) + (n_2/\sigma_2^2)},$$

so that

$$\text{var}(\hat{\mu}_O) = \frac{1}{(n_1/\sigma_1^2) + (n_2/\sigma_2^2)}.$$

Then

$$\text{(i)}\quad \text{var}(\hat{\mu}^{(2)}) = \frac{m_2(m_2+2)}{(m_1+m_2)(m_1+m_2+2)}\ {}_2F_1\left[2, \frac{m_1}{2}; \frac{m_1+m_2+4}{2}; \frac{m_1-m_2c}{m_1}\right].$$

$$\left(\frac{\sigma_1^2}{n_1} + c^2 m_2 \frac{\sigma_2^2}{n_2}\right), \qquad (3.25)$$

where $m_1 = n_1 - 1$, $m_2 = n_2 - 1$, $c = \dfrac{\alpha_2\sigma_1^2}{\alpha_1\sigma_2^2}$ and ${}_2F_1(\cdot,\cdot;\cdot;\cdot)$ is the hypergeometric function.

$$\text{(ii)}\quad \frac{\text{var}(\hat{\mu}^{(2)})}{\text{var}(\hat{\mu}_O)} \le \frac{1}{2} + \frac{1}{4}\left[\frac{\alpha_1 n_2(n_1-1)}{\alpha_2 n_1(n_1-3)} + \frac{\alpha_2 n_1(n_2-1)}{\alpha_1 n_2(n_2-3)}\right]. \qquad (3.26)$$

Proof: (i)

Consider (3.16) for the special case of $r=2$. i.e.,

$$\text{var}(\hat{\mu}^{(2)}) = \frac{\sigma_1^2}{n_1} E\left[\frac{1}{(1+cF)^2}\right] + \frac{\sigma_2^2}{n_2} E\left[\frac{c^2F^2}{(1+cF)^2}\right]. \qquad (3.27)$$

Now if we let $X = \dfrac{1}{(1+cF)}$ then recalling that $F \sim F_{n_1-1, n_2-1}$ and using transformation of variables one can show that the probability density function of X is given by

$$f_X(x) = \frac{m_1^{m_1/2} m_2^{m_2/2} c^{m_2/2}}{B(m_1/2, m_2/2)} \frac{x^{(m_2-2)/2}(1-x)^{(m_1-2)/2}}{[m_1 + (m_2 c - m_1)x]^{(m_1+m_2)/2}},$$

where $0 \le x \le 1$ and $B(\cdot,\cdot)$ is the beta function.

Hence

$$E\left[\frac{1}{(1+cF)^2}\right] = K \int_0^1 \frac{x^{(m_2+2)/2}(1-x)^{(m_1-2)/2}}{[m_1 + (m_2 c - m_1)x]^{(m_1+m_2)/2}} dx$$

$$= \frac{K}{m_1^{(m_1+m_2)/2}} \int_0^1 \frac{x^{(m_2+2)/2}(1-x)^{(m_1-2)/2}}{[1 + (\frac{m_2 c - m_1}{m_1})x]^{(m_1+m_2)/2}} dx, \qquad (3.28)$$

where

$$K = \frac{m_1^{m_1/2} m_2^{m_2/2} c^{m_2/2}}{B(m_1/2, m_2/2)}.$$

Now by the integral representation of the hypergeometric function (see Bell, 1968, p. 207) we obtain

$$\int_0^1 \frac{x^{(m_2+2)/2}(1-x)^{(m_1-2)/2}}{[1 + (\frac{m_2 c - m_1}{m_1})x]^{(m_1+m_2)/2}} dx$$

$$= {}_2F_1\left[\frac{m_1+m_2}{2}, \frac{m_2+4}{2}; \frac{m_1+m_2+4}{2}; \frac{m_1 - m_2 c}{m_1}\right]. \qquad (3.29)$$

Substituting (3.29) in (3.28), further simplifying the beta function and using identities concerning hypergeometric functions (see Bell, 1968, p. 208) we obtain

$$E\left[\frac{1}{(1+cF)^2}\right] = \frac{m_2(m_2+2)}{(m_1+m_2)(m_1+m_2+2)} \; {}_2F_1\left[2, \frac{m_1}{2}; \frac{m_1+m_2+4}{2}; \frac{m_1 - m_2 c}{m_1}\right]. \quad (3.30)$$

Further observe that

$$\frac{c^2F^2}{(1+cF)^2} = \frac{1}{(1+c^*F^*)^2},$$

where $c^* = \frac{1}{c}$ and $F^* \sim F_{m_2,m_1}$.

By a similar consideration as before we obtain

$$E\left[\frac{c^2F^2}{(1+cF)^2}\right] = \frac{m_2^2(m_2+2)}{(m_1+m_2)(m_1+m_2+2)}c^2\ {}_2F_1\left[2,\frac{m_1}{2};\frac{m_1+m_2+4}{2};\frac{m_1-m_2c}{m_1}\right]. \tag{3.31}$$

Now by substituting (3.30) and (3.31) in (3.27) we obtain the required result.

Proof: (ii)

First we observe $\mathrm{var}(\hat{\mu}^{(2)})$ depends on S_1^2 and S_2^2, only through the ratio S_1^2/S_2^2 and $\hat{\mu}^{(2)}$ is an unbiased estimator of the common mean μ. Now

$$\begin{aligned}
\mathrm{var}(\hat{\mu}^{(2)}) &= E\left(\hat{\mu}^{(2)^2}\right) - \left(E(\hat{\mu}^{(2)})\right)^2 \\
&= E\left(E\left(\hat{\mu}^{(2)^2} \mid S_1^2/S_2^2\right)\right) - \mu^2 \\
&= E\left(E\left(\hat{\mu}^{(2)^2} \mid S_1^2/S_2^2\right) - \mu^2\right) \\
&= E\left(E\left(\hat{\mu}^{(2)^2} \mid S_1^2/S_2^2\right) - \left(E\left(\hat{\mu}^{(2)} \mid S_1^2/S_2^2\right)\right)^2\right) \\
&= E\left(\mathrm{var}(\hat{\mu}^{(2)} \mid S_1^2/S_2^2)\right).
\end{aligned} \tag{3.32}$$

Let $\hat{\mu}_W = w_1\bar{Y}_1 + w_2\bar{Y}_2$ where $w_1 + w_2 = 1$. Then

$$\begin{aligned}
\mathrm{var}(\hat{\mu}_W \mid S_1^2/S_2^2) &= \mathrm{var}(w_1\bar{Y}_1 + w_2\bar{Y}_2 \mid S_1^2/S_2^2) \\
&= \mathrm{var}(w_1\bar{Y}_1 + w_2\bar{Y}_2).
\end{aligned}$$

Therefore

$$\min_{w_1, w_2} \operatorname{var}(\hat{\mu}_W | S_1^2/S_2^2) = \operatorname{var}(\hat{\mu}_O).$$

Now using Theorem 3.3.1, we can write

$$\frac{\operatorname{var}(\hat{\mu}^{(2)} | S_1^2/S_2^2)}{\operatorname{var}(\hat{\mu}_O)} \le \frac{(R+1)^2}{4R} = \frac{1}{2} + \frac{1}{4}(R + \frac{1}{R}), \tag{3.33}$$

where

$$R = \frac{\max\{(a_1/S_1^2)(\sigma_1^2/n_1), (a_2/S_2^2)(\sigma_2^2/n_2)\}}{\min\{(a_1/S_1^2)(\sigma_1^2/n_1), (a_2/S_2^2)(\sigma_2^2/n_2)\}}.$$

It is a simple matter to observe that, no matter what the ratio of S_1^2/S_2^2 is

$$R + \frac{1}{R} = \frac{a_1 n_2}{a_2 n_1} \frac{S_2^2/\sigma_2^2}{S_1^2/\sigma_1^2} + \frac{a_2 n_1}{a_1 n_2} \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2},$$

and

$$\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F_{n_1-1, n_2-1} \quad \text{and} \quad \frac{S_2^2/\sigma_2^2}{S_1^2/\sigma_1^2} \sim F_{n_2-1, n_1-1}.$$

Therefore by taking expectations of both sides of (3.33) we obtain

$$\frac{\operatorname{var}(\hat{\mu}^{(2)})}{\operatorname{var}(\hat{\mu}_O)} \le \frac{1}{2} + \frac{1}{4}\left[\frac{a_1 n_2 (n_1 - 1)}{a_2 n_1 (n_1 - 3)} + \frac{a_2 n_1 (n_2 - 1)}{a_1 n_2 (n_2 - 3)}\right]. \tag{3.34}$$

Q·E·D

It is interesting to notice here that we can minimize the above upper bound for inefficiency of $\operatorname{var}(\hat{\mu}^{(2)})$ by choosing

$$\frac{a_1}{a_2} = \frac{n_1}{n_2}\left[\frac{(n_1 - 3)(n_2 - 1)}{(n_1 - 1)(n_2 - 3)}\right]^{1/2}.$$

If we substitute the above choice of $\frac{a_1}{a_2}$ in (3.34) we obtain

$$\frac{\operatorname{var}(\hat{\mu}^{(2)})}{\operatorname{var}(\hat{\mu}_o)} \leq \frac{1}{2}+\frac{1}{2}\left[\frac{(n_1-1)(n_2-1)}{(n_1-3)(n_2-3)}\right]^{1/2} = U(n_1,n_2) \quad \text{(say)}. \tag{3.35}$$

Contour plots of U for different values of n_1 and n_2 are given in Figure 3.1. A quick inspection of Figure 3.1 shows that with the above choice of $\frac{a_1}{a_2}$ and for samples of size greater than 15, inefficiency of $\hat{\mu}^{(2)}$ is less than or equal to 1.09.

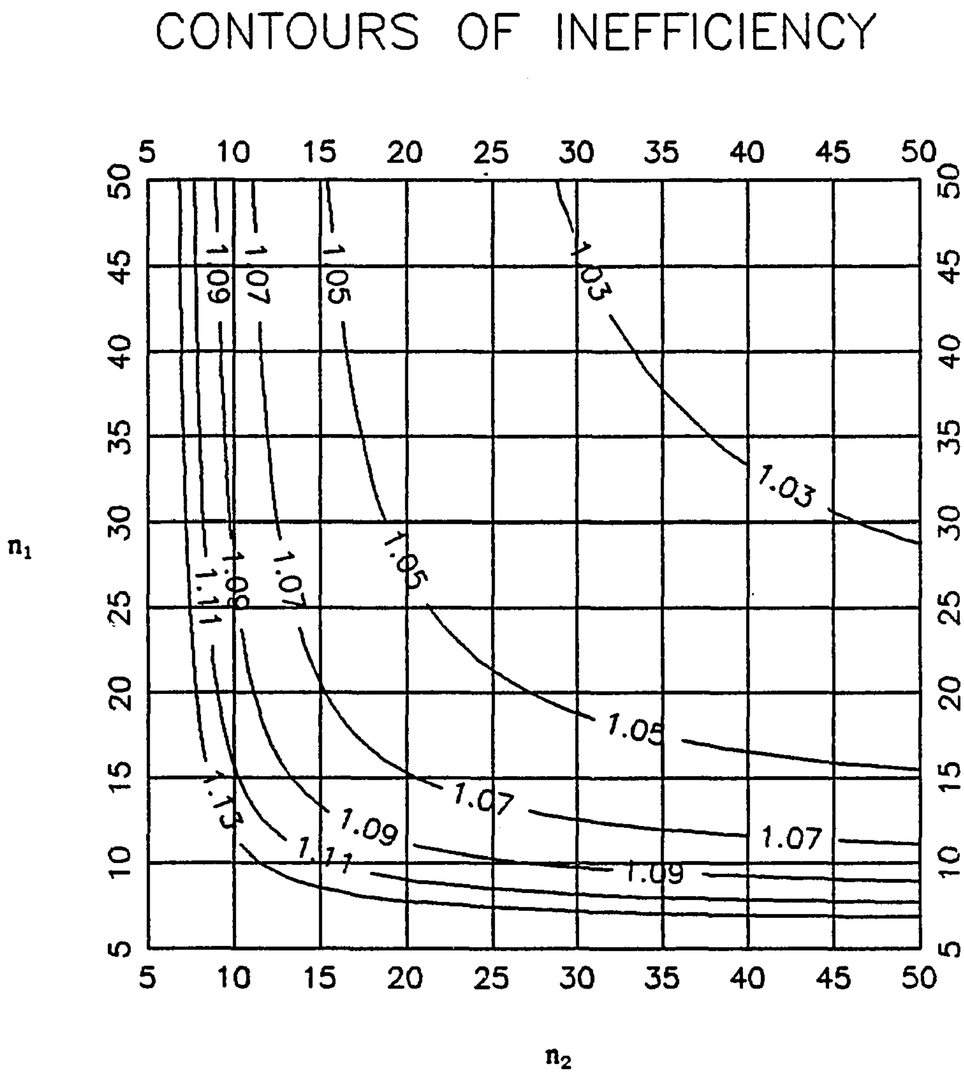


Figure 3.1. Contours of inefficiency.

3.4. Safe T—statistics (Cressie, 1982)

In Section 3.2 we discussed the importance of weighted estimation of a common mean μ when the observations are heteroskedastic, while in Section 3.3 we presented results on weighted estimation with both deterministic and random weights. Most often we are interested in hypothesis testing problems concerning the common mean μ. In this section we will present the notion of "safeness" and show how to construct a safe T—statistic (when a weighted mean is available to us as a point estimate of the common mean μ) to make inference about the common mean, even when the weights are misspecified. Asymptotic distribution of this T—statistic and finite sample considerations will also be discussed.

<u>3.4.1</u> <u>Definition</u> (Cressie, 1982)

Let $\hat{\theta}_n$ be an estimator of an unknown population parameter θ and let a_n be an estimator of $\mathrm{var}(\hat{\theta}_n)$. Let $T = \dfrac{\hat{\theta}_n - \theta}{\sqrt{a_n}}$. Then T is called a *safe test statistic* if $E(a_n) = \mathrm{var}(\hat{\theta}_n)$, i.e., if a_n is an unbiased estimator of $\mathrm{var}(\hat{\theta}_n)$ and T is called *asymptotically safe* if $a_n \rightarrow \lim_{n\rightarrow\infty} \mathrm{var}(\hat{\theta}_n)$ in pr, if this limiting variance exists otherwise, it is *asymptotically safe* if $|a_n - \mathrm{var}(\hat{\theta}_n)| \rightarrow 0$, in pr.

Let $Y_1, Y_2, \ldots, Y_n$ be independent observations with common mean μ and variances (possibly different) σ_i^2; $i = 1,2,\ldots,n$, respectively. Let

$$\bar{Y}_w = \sum_{i=1}^{n} w_i Y_i \; ; \quad \sum_{i=1}^{n} w_i = 1, \tag{3.36}$$

$$S^2_{w,\delta} = \sum_{i=1}^{n} \delta_i (Y_i - \bar{Y}_w)^2; \;\; \delta_i > 0, \; i = 1,2,\ldots,n, \tag{3.37}$$

and

$$T_W = \frac{\bar{Y}_W - \mu}{S_{w,\delta}}. \tag{3.38}$$

We assume that the weights $\{w_i\}$ and $\{\delta_i\}$ are fixed deterministic quantities.

Cressie (1982) shows that even if the weights $\{w_i\}$ in (3.36) are *misspecified*, there always exists a set of compensating weights $\{\delta_i\}$ in the definition of $S^2_{w,\delta}$ given in (3.37), that makes the test statistic T_W defined by (3.38) *safe*. We will go through his derivations below.

For safeness of the test statistic T_W, we need to choose δ_i such that

$$E(S^2_{w,\delta}) = \mathrm{var}(\bar{Y}_W).$$

Since the weights $\{w_i\}$ are fixed we obtain

$$\mathrm{var}(\bar{Y}_W) = \sum_{i=1}^{n} w_i^2 \sigma_i^2 ,$$

and

$$\begin{aligned} E(S^2_{w,\delta}) &= \sum_{i=1}^{n} \delta_i E(Y_i - \bar{Y}_W)^2 \\ &= \sum_{i=1}^{n} \delta_i [\sigma_i^2 (1 - w_i)^2 + \sum_{j \neq i}^{n} w_j^2 \sigma_j^2] \\ &= \sum_{i=1}^{n} [w_i^2 \delta_+ - 2 w_i \delta_i + \delta_i] \sigma_i^2, \end{aligned}$$

where $\delta_+ = \delta_1 + \delta_2 + \ldots + \delta_n$.

Now matching coefficients of σ_i^2, we obtain

$$\delta_i = \frac{w_i^2(1-\delta_+)}{(1-2w_i)} \quad (i=1,2,\ldots,n). \tag{3.39}$$

Summing (3.39) over $i=1,2,\ldots,n$, we obtain

$$\delta_+ = (1-\delta_+)\sum_{i=1}^{n}\frac{w_i^2}{(1-2w_i)}.$$

Therefore

$$\delta_+ = \frac{\sum_{i=1}^{n} w_i^2(1-2w_i)^{-1}}{\{1+\sum_{i=1}^{n} w_i^2(1-2w_i)^{-1}\}}, \tag{3.40}$$

and

$$\delta_i = \frac{w_i^2(1-2w_i)^{-1}}{\{1+\sum_{i=1}^{n} w_i^2(1-2w_i)^{-1}\}}. \tag{3.41}$$

Hence we see that for any set of fixed weights $\{w_i\}$ there exists a set of compensating weights $\{\delta_i\}$ that makes T_W safe. We easily notice that when we use equal weights for w_i (i.e., $w_i = \frac{1}{n}$) then $\delta_i = \frac{1}{n(n-1)}$ and thus we obtain the usual T—statistic.

Now consider T_W given by (3.38) where $\{\delta_i\}$ are chosen such that (3.41) holds. Then T_W is a *safe* test statistic. Cressie (1982) obtains the asymptotic distribution of T_W assuming that the n independent observations can be divided into p strata so that equal variation occurs within each stratum; i.e.,

$$\frac{Y_{ij}-\mu}{\sigma_j} \sim G \qquad (j=1,2,\ldots,n_i,\ i=1,2,\ldots,p), \tag{3.42}$$

and

$$\frac{n_i}{n} \rightarrow \theta_i \,, \quad \text{as } n \rightarrow \infty, \tag{3.43}$$

where $\sum_{i=1}^{p} n_i = n$, $0 < \theta_i < 1$, and $\sum_{i=1}^{p} \theta_i = 1$.

3.4.2. Theorem (Cressie, 1982)

Under the Assumptions (3.42) and (3.43), $T_W \xrightarrow{d} Z$, where Z is the standard normal random variable.

We refer the reader to Cressie (1982) for the proof of this theorem. The theorem above allows us to construct confidence intervals for the common mean μ and perform tests concerning μ, at least asymptotically, even when the weights $\{w_i\}$ are misspecified.

The next question to ask is, "Can the finite sample distribution of T_W be approximated by a t—distribution with some equivalent degrees of freedom?

In his paper Cressie (1982) argues and shows that under an assumption of normality of the observations, T_W can be approximated by a t—distribution with equivalent degrees of freedom

$$\text{e.d.f} = \left\{ \sum_{i=1}^{n} \lambda_i^2 \right\}^{-1}, \text{ to } O(1), \tag{3.44}$$

where

$$\lambda_i = \frac{\delta_i \tau_i^2}{\sum_{j=1}^{n} \delta_j \tau_j^2} \qquad (i = 1, 2, \ldots, n),$$

and

$$\tau_j^2 = \mathrm{var}(Y_j - \bar{Y}_W).$$

To examine the appropriateness of the equivalent degrees of freedom, a simulation study was carried out for the special case of $p = 3$ (i.e., three strata) where the equal variation within each stratum was taken to be $\sigma_1^2 = 1$, $\sigma_2^2 = 9$, and $\sigma_3^2 = 81$ respectively. In this simulation we considered a situation where an equal number of observations occurred in each stratum; i.e., $n_1 = n_2 = n_3$. The number of observations that came from each stratum were taken as 4, 10, and 20. Data were generated using the IMSL double precision normal random number generator DRNNOR. We considered three weighting schemes in this simulation, namely, equal weights which give rise to the usual T—statistic, weights proportional to individual sample standard deviations $(1/S_i)$, and weights proportional to individual sample variances $(1/S_i^2)$. Even though the results given by Cressie (1982) are proved for fixed weights, it is interesting to see how T_W performs when random weights depending on S_i^2, are chosen. With these weighting schemes we obtain three T—statistics. The following names were given to the different T—statistics for identification purposes.

T : corresponds to equal weights. (i.e., the usual T—statistic),

TNCS : corresponds to the weights proportional to $1/S_i$,

TNCS2 : corresponds to the weights proportional to $1/S_i^2$.

When the observations are homoskedastic T follows a t—distribution with $(n-1)$ degrees of freedom. If we relax the homoskedasticity assumption even with equal weights one might approximate the distribution of T by a t—distribution with equivalent degrees of freedom given by (3.44). Also we calculated the equivalent degrees of freedom corresponding to TNCS and TNCS2 using the formula (3.44). The simulation was replicated 1000 times. Since each replication provided an e.d.f. for TNCS and TNCS2 we considered the harmonic means of the equivalent degrees of freedom for the final Q—Q plots. We used IMSL double precision subroutine DTIN to obtain the approximate expected values of order statistics from Student's t—distribution. Figures 3.2—3.13 show the resulting Q—Q plots. These Q—Q plots clearly indicate that the weights proportional to $1/S_i$ are superior to the weights proportional to $1/S_i^2$ and to the equal weights case, especially for small sample sizes. For large samples (i.e., $n_i \geq 20$) the usual T with $(n-1)$ d.f. or e.d.f. seems to approximate the distribution by a t—distribution fairly closely. Thus we would recommend for further investigation, using weights proportional to $1/S_i$, provided the groups of unequal variation are known. Otherwise, for $n_i \geq 20$, it appears that the usual T—statistic gives valid inference.

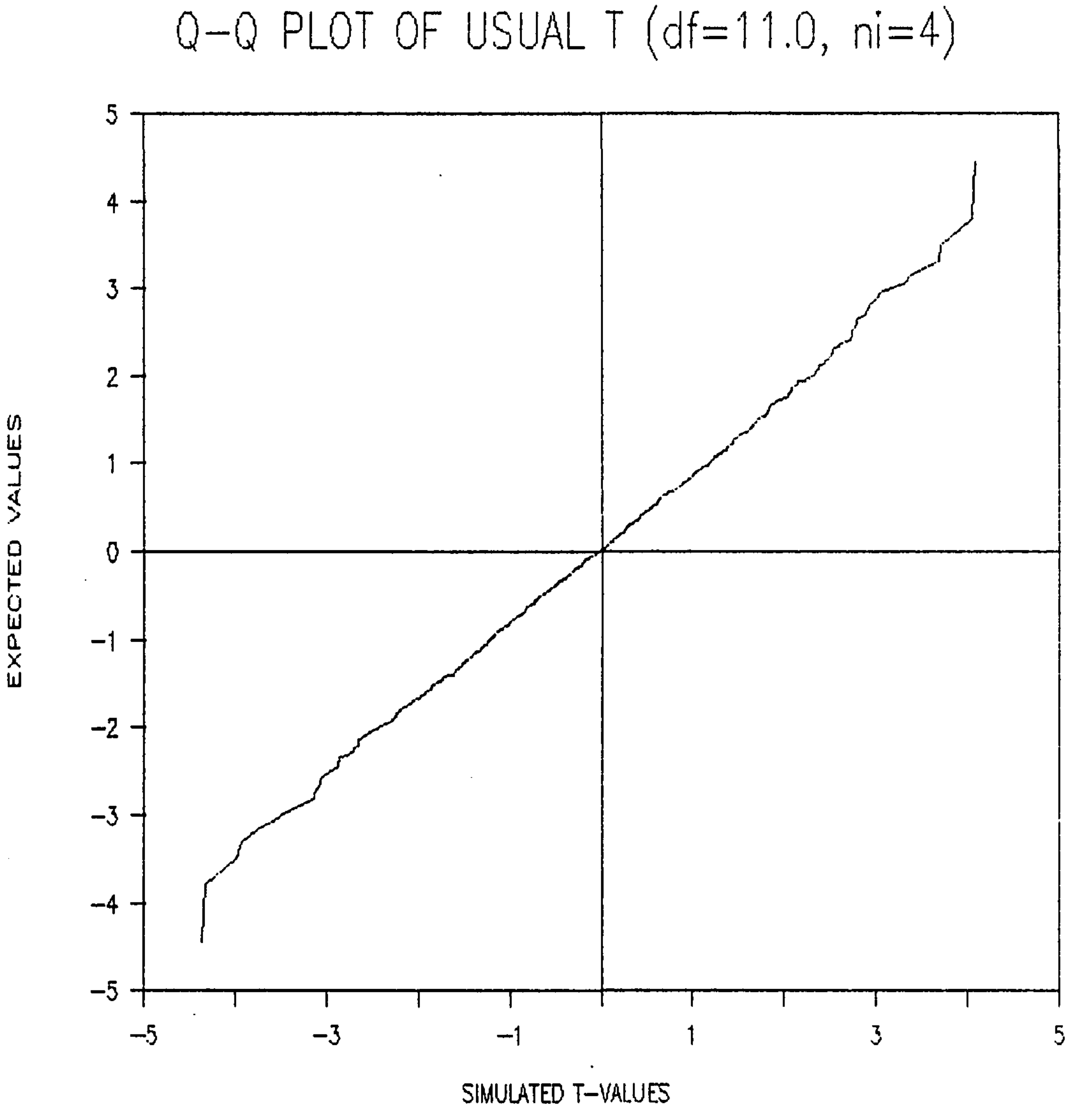


Figure 3.2. Q—Q plot of usual T with usual degrees of freedom, $n_1 = n_2 = n_3 = 4$

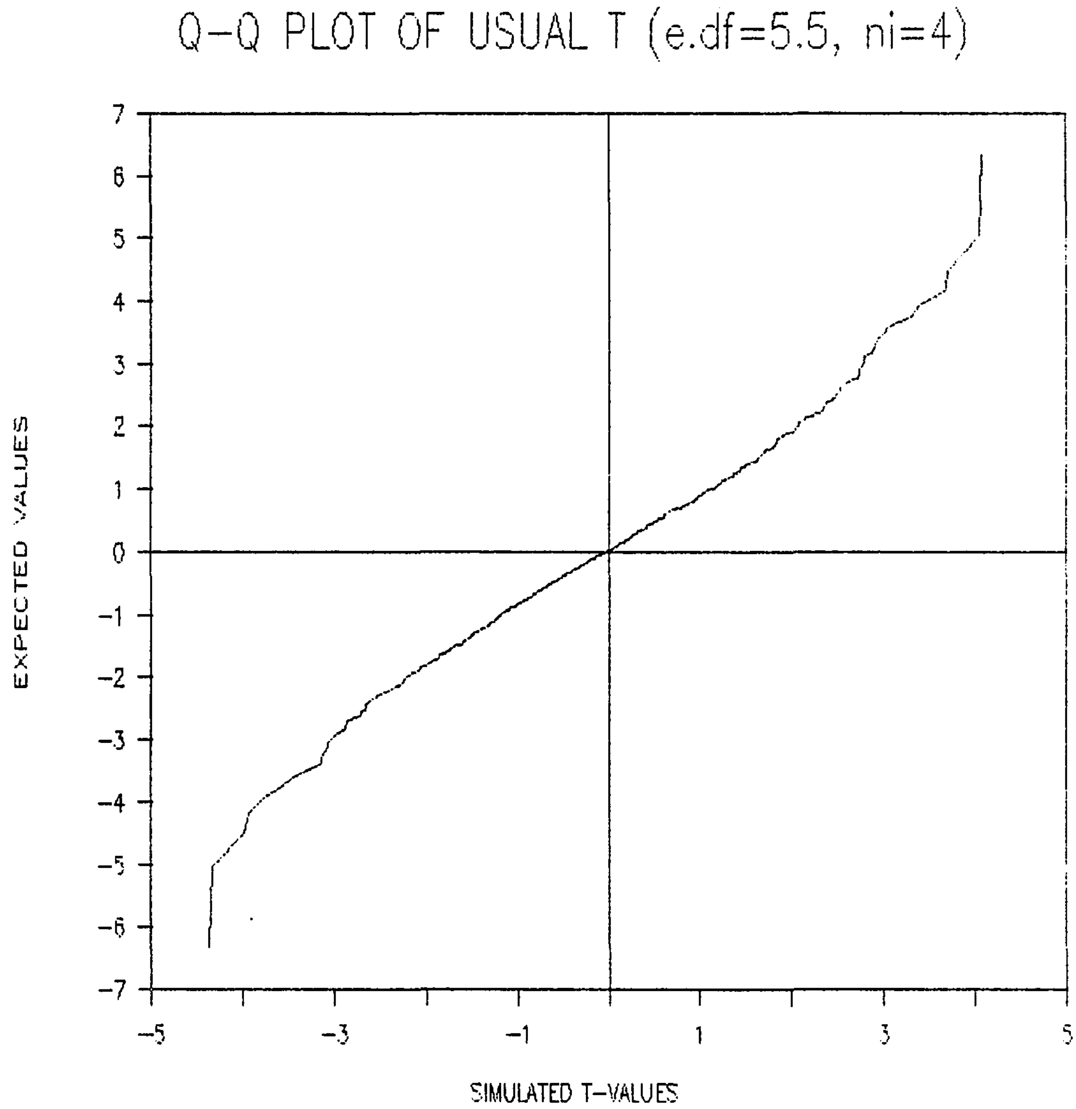


Figure 3.3. Q—Q plot of usual T with equivalent degrees of freedom, $n_1 = n_2 = n_3 = 4$

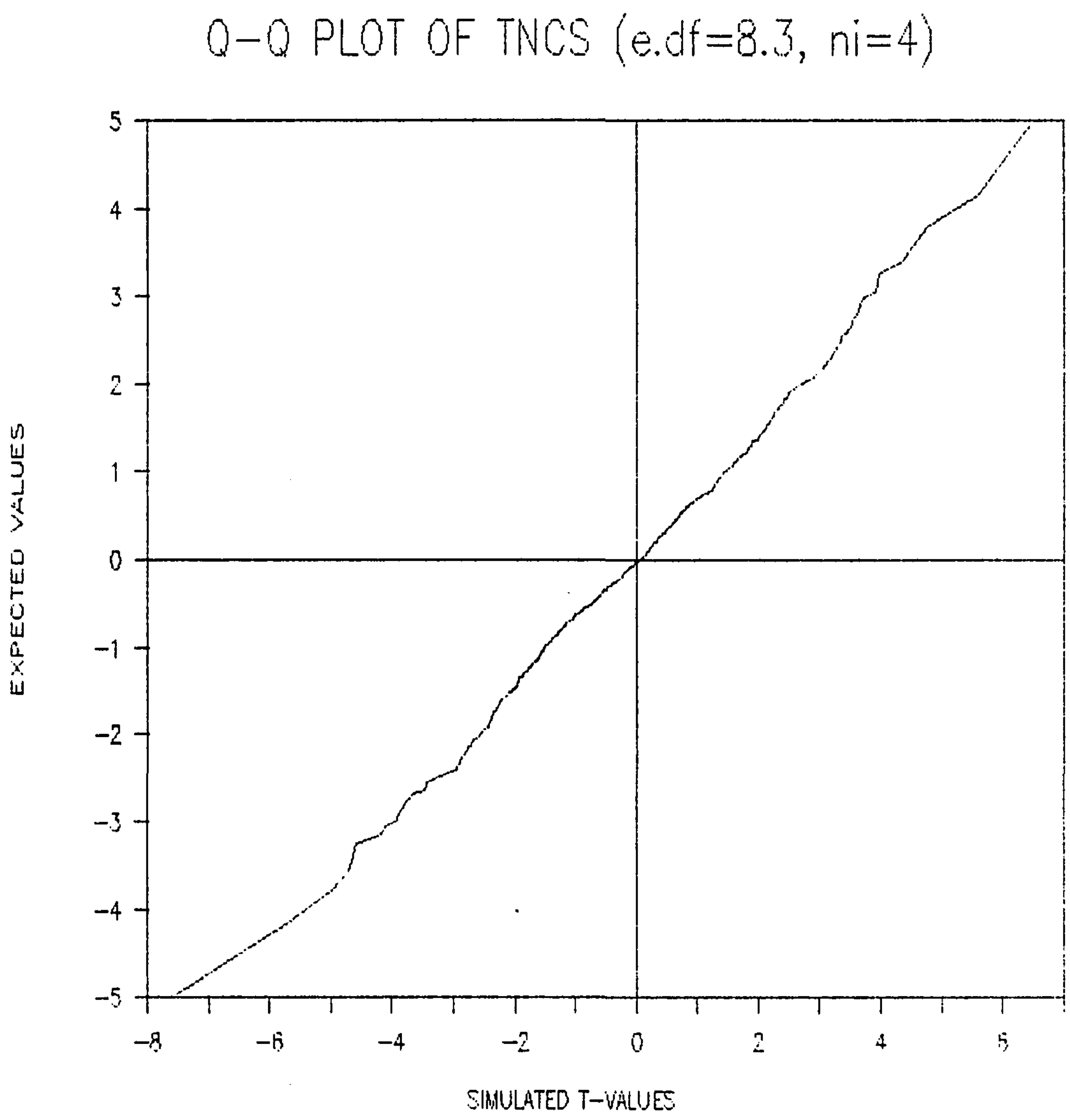


Figure 3.4. Q—Q plot of TNCS, i.e., weights $\propto 1/S_i$, $n_1 = n_2 = n_3 = 4$

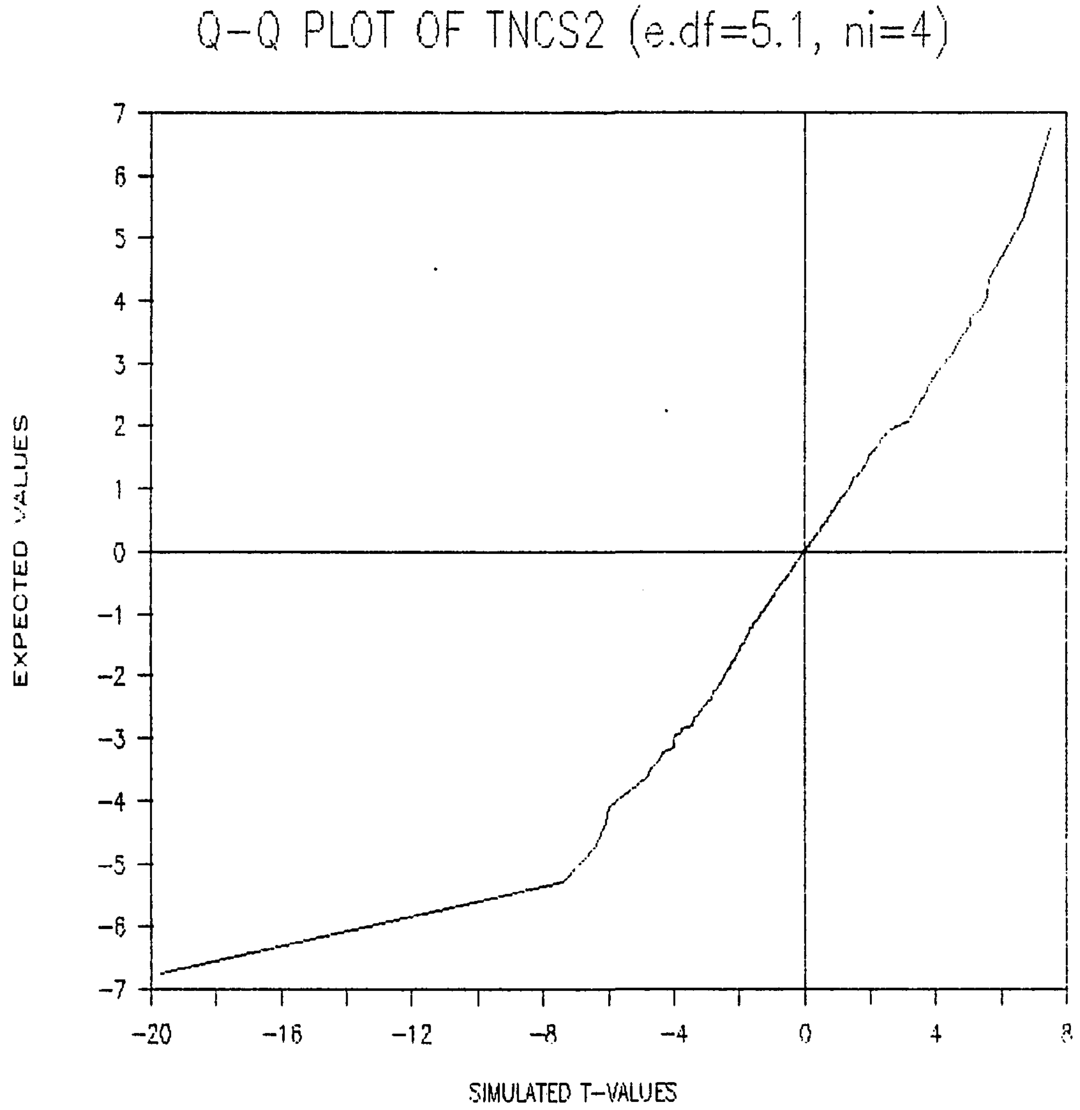


Figure 3.5. Q—Q plot of TNCS2, i.e., weights $\propto 1/S_i^2$, $n_1 = n_2 = n_3 = 4$

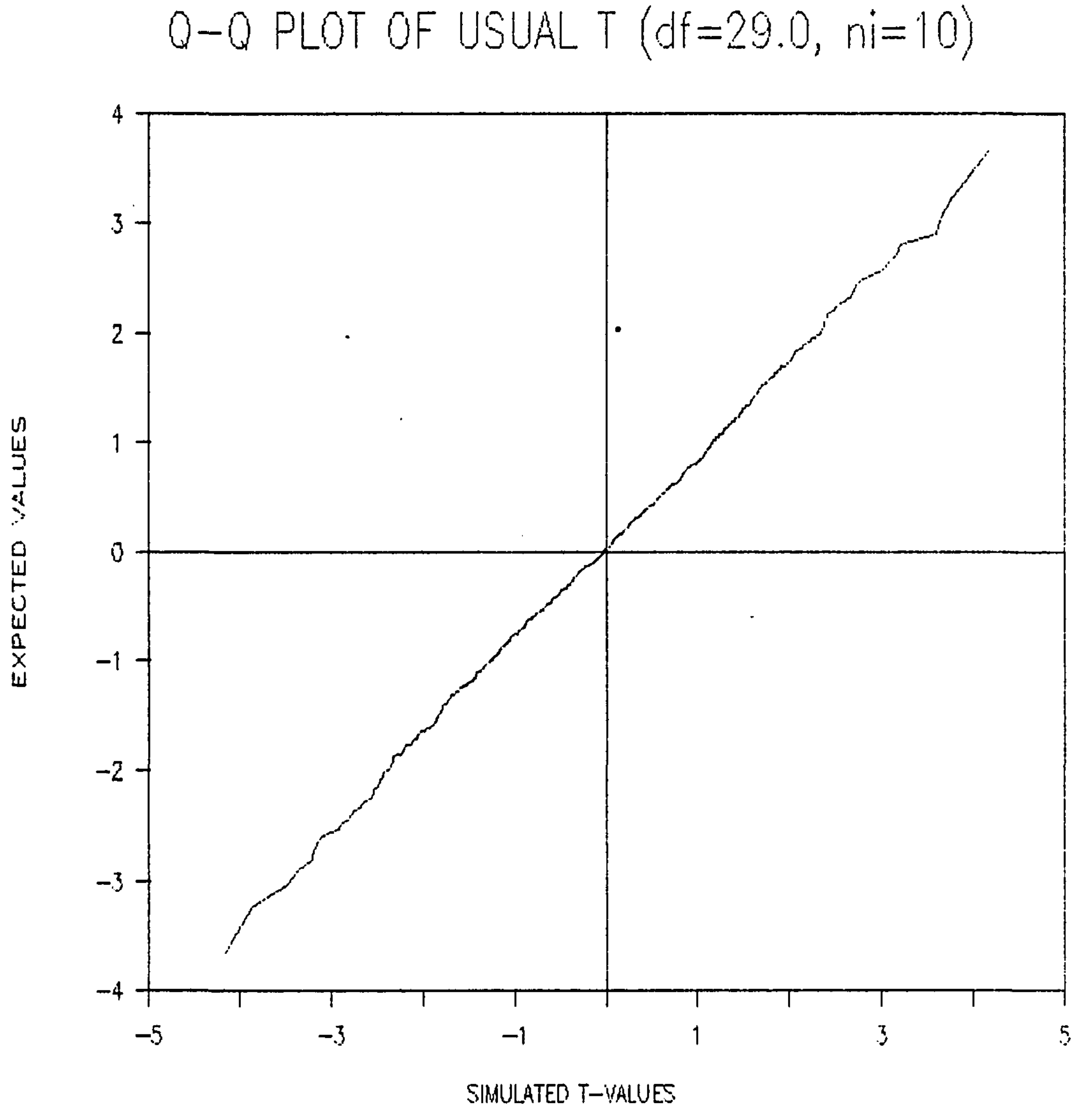


Figure 3.6. Q–Q plot of usual T with usual degrees of freedom, $n_1 = n_2 = n_3 = 10$

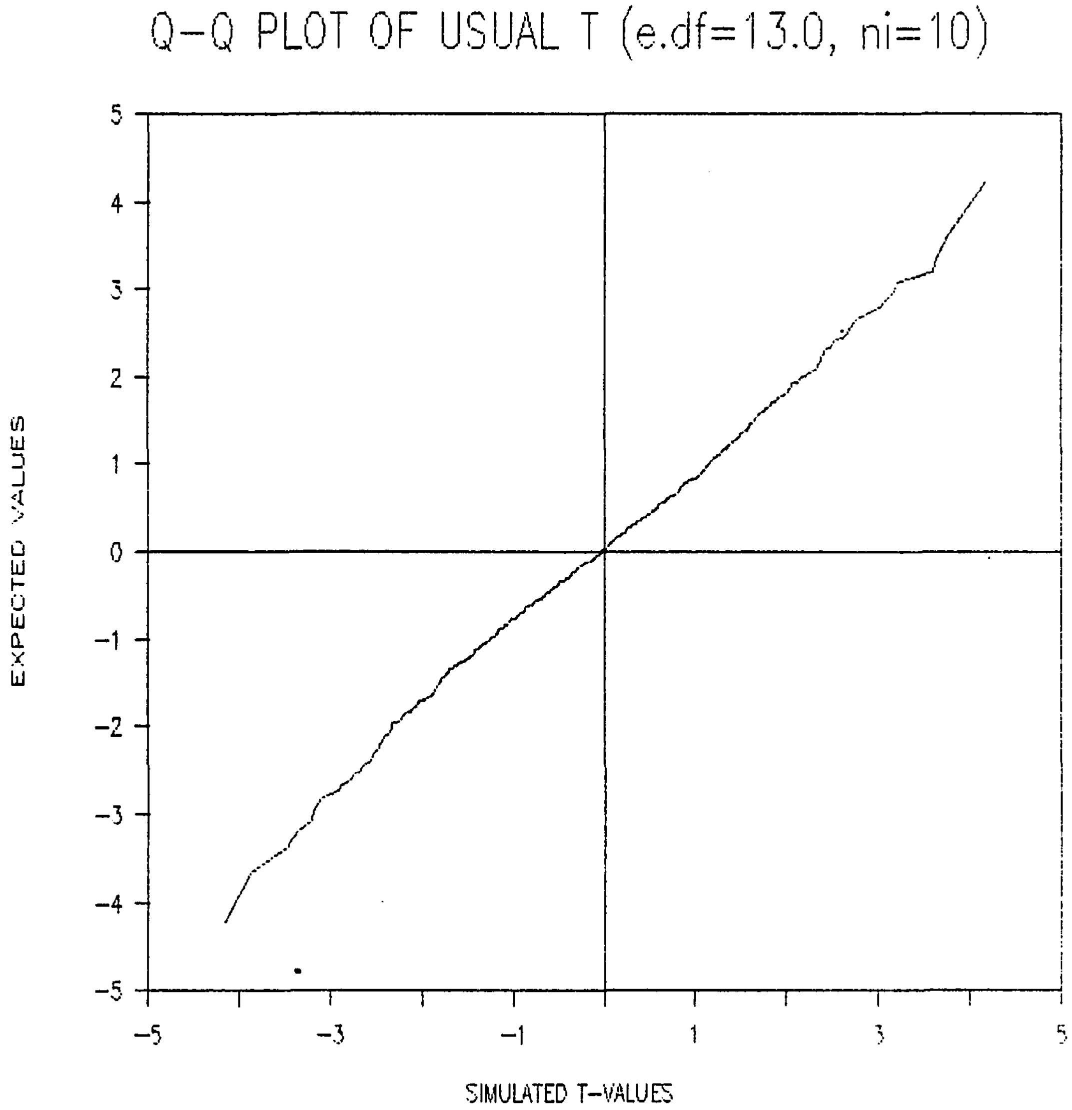


Figure 3.7. Q—Q plot of usual T with equivalent degrees of freedom, $n_1 = n_2 = n_3 = 10$

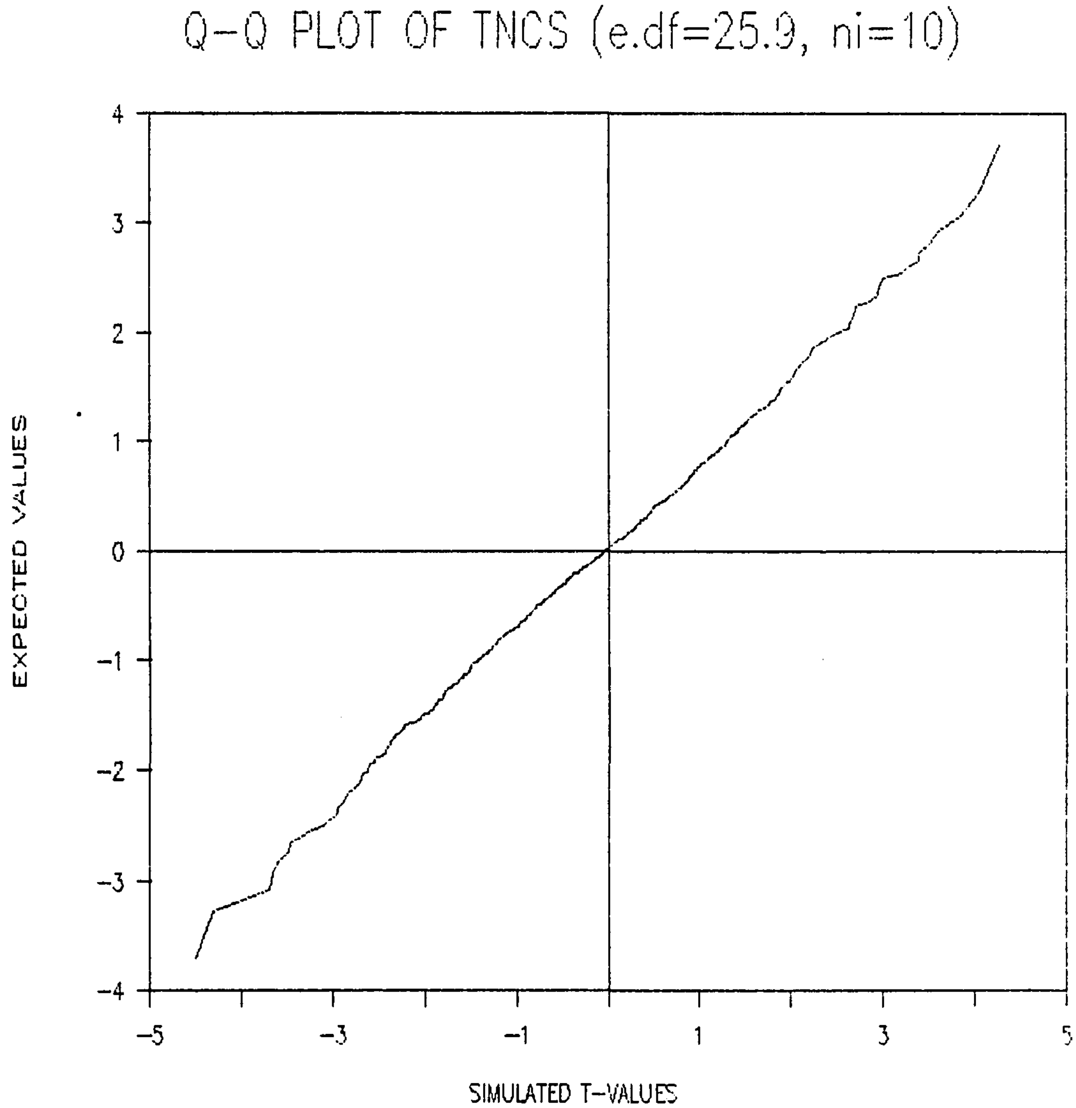


Figure 3.8. Q—Q plot of TNCS, i.e., weights $\propto 1/S_i$, $n_1 = n_2 = n_3 = 10$

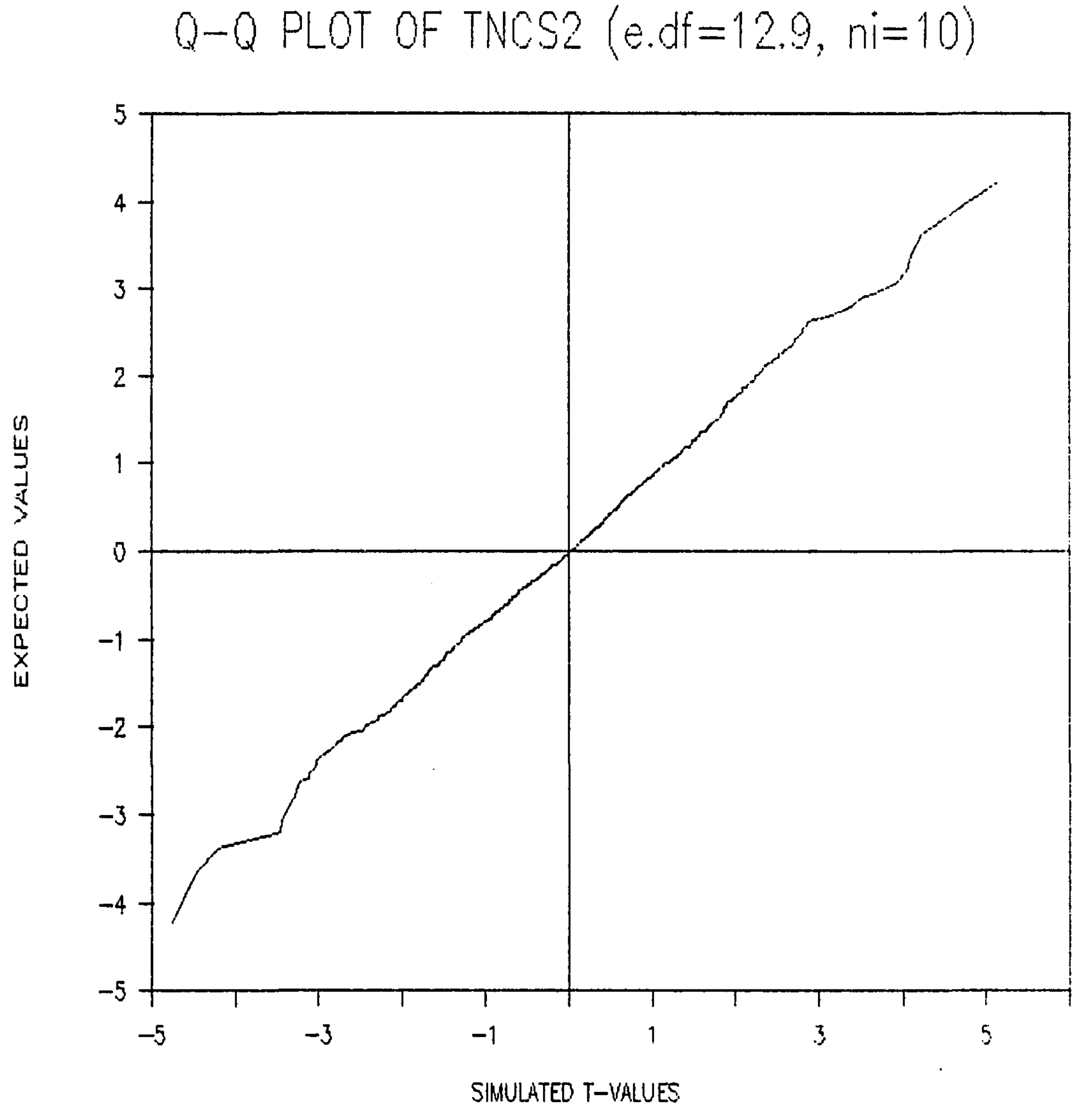


Figure 3.9. Q—Q plot of TNCS2, i.e., weights $\propto 1/S_i^2$, $n_1 = n_2 = n_3 = 10$

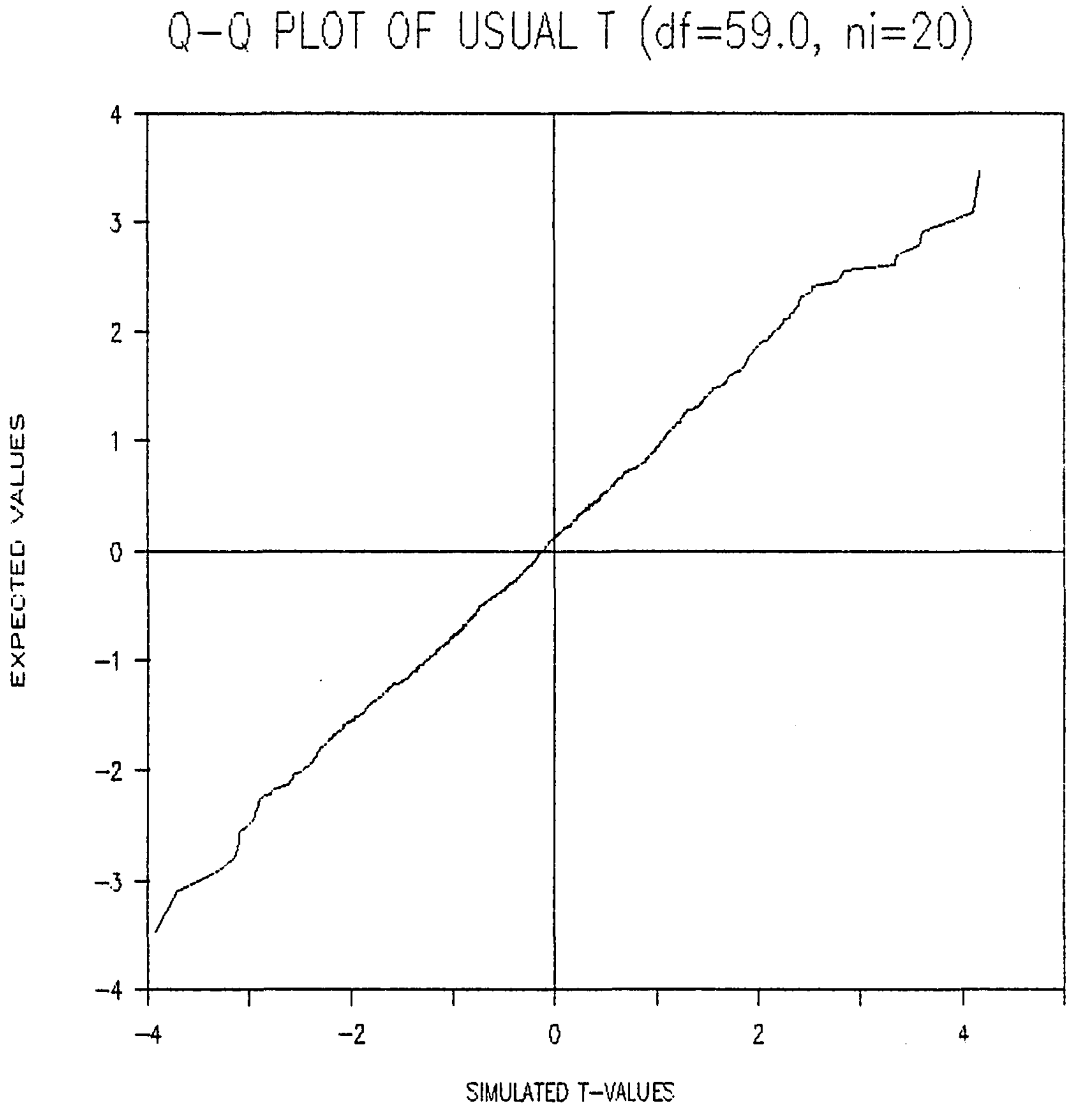


Figure 3.10. Q—Q plot of usual T with usual degrees of freedom, $n_1 = n_2 = n_3 = 20$

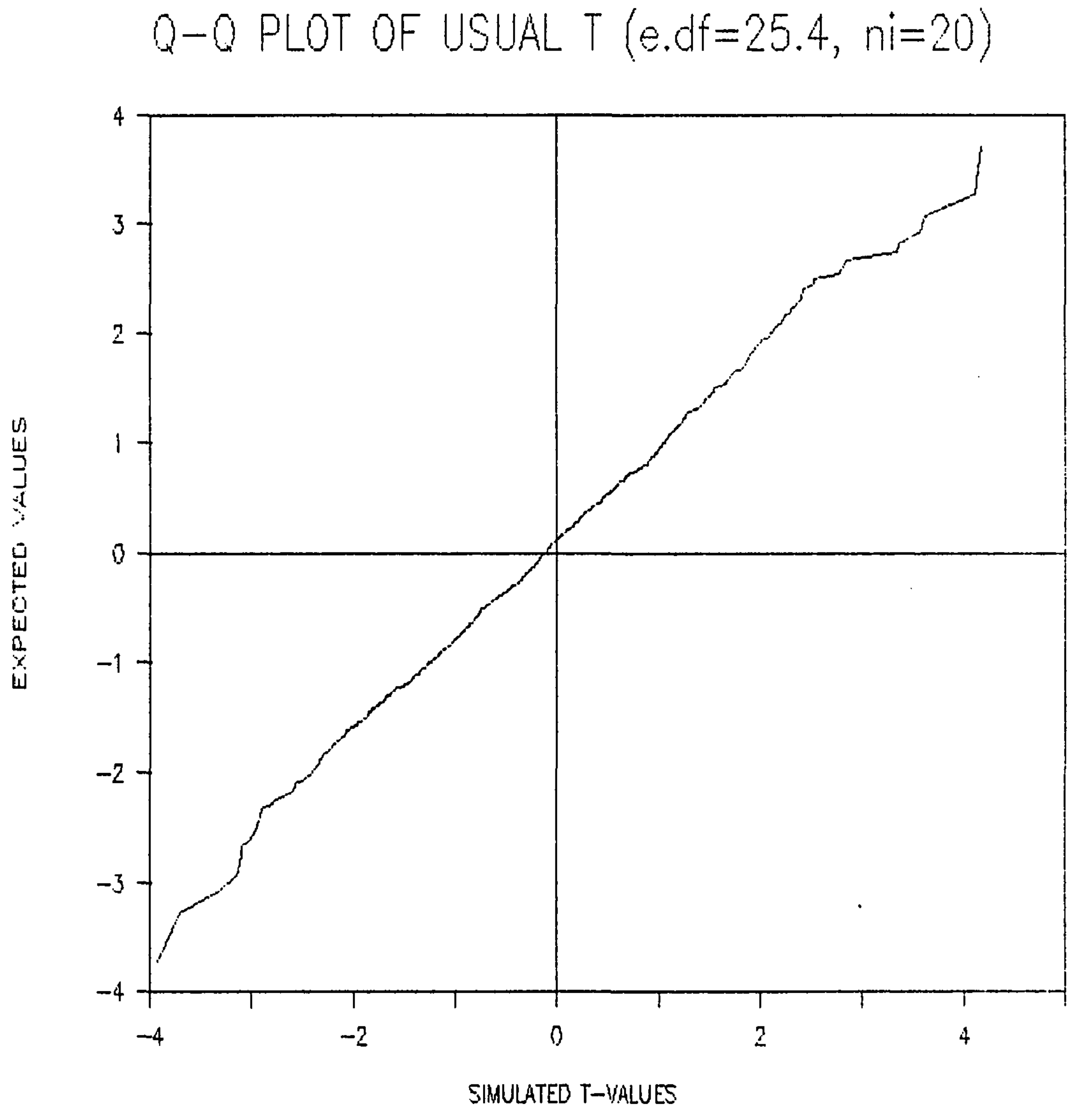


Figure 3.11. Q—Q plot of usual T with equivalent degrees of freedom, $n_1 = n_2 = n_3 = 20$

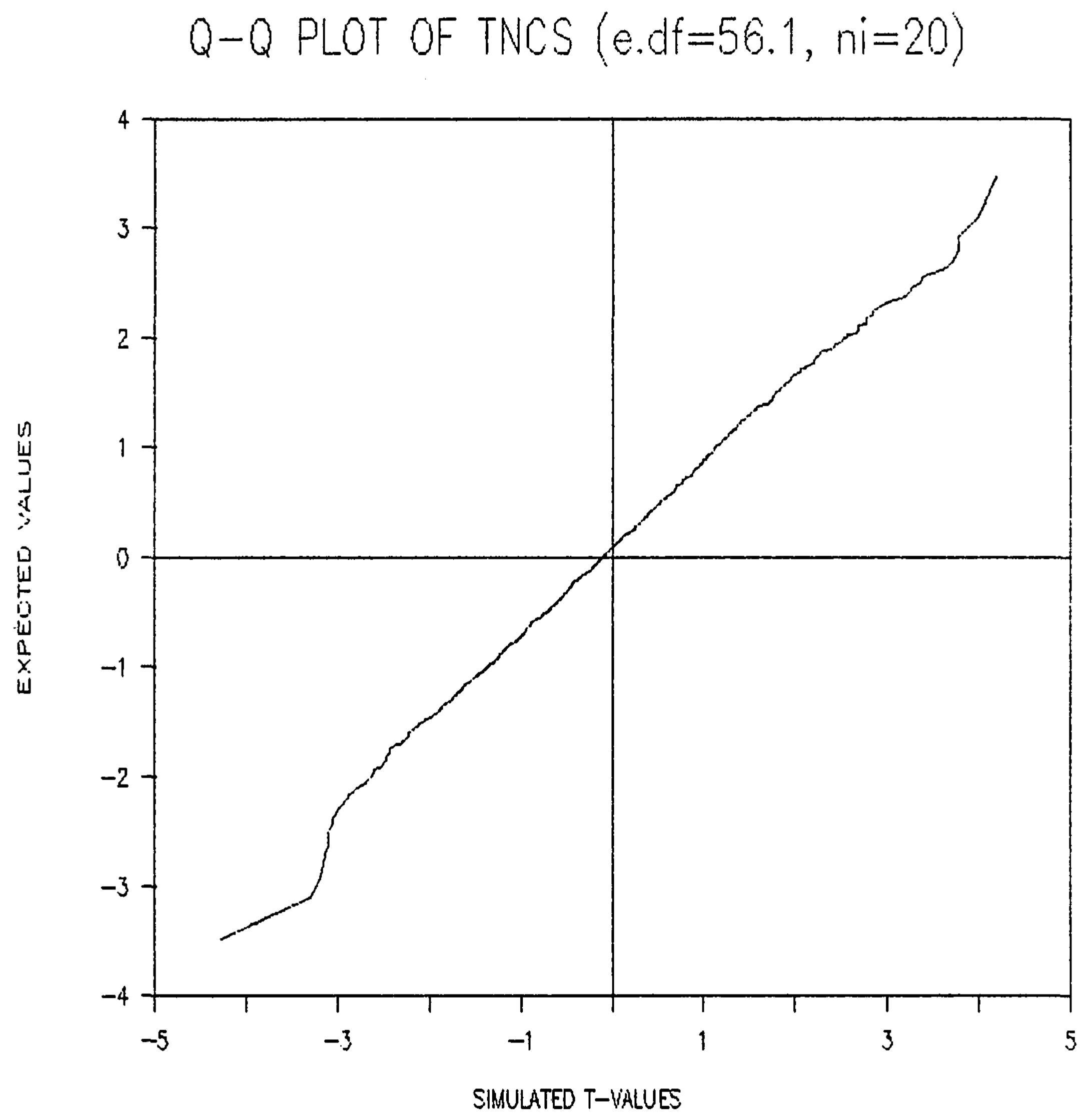


Figure 3.12. Q–Q plot of TNCS, i.e., weights $\propto 1/S_i$, $n_1 = n_2 = n_3 = 20$

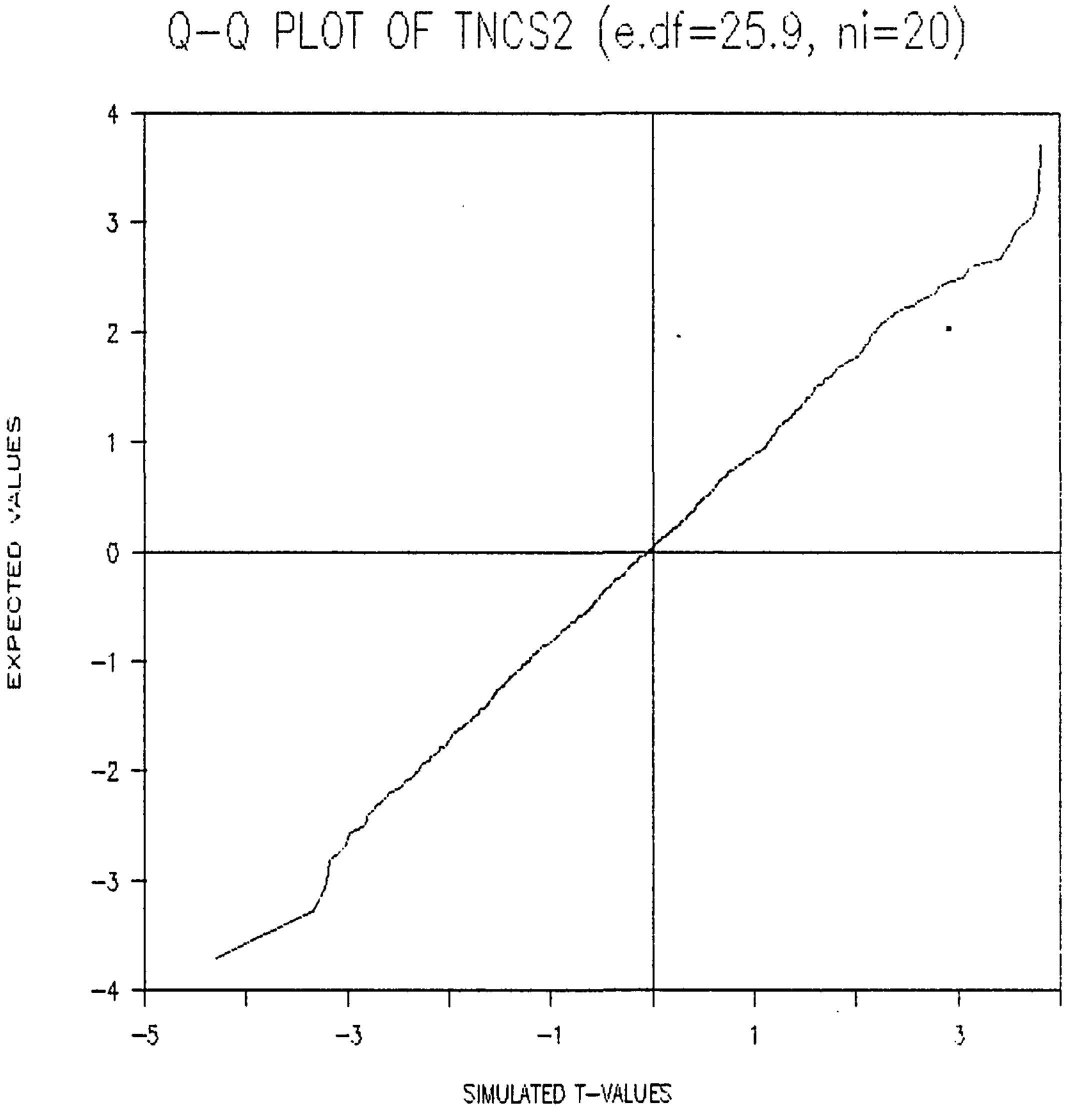


Figure 3.13. Q—Q plot of TNCS2, i.e., weights $\propto 1/S_i^2$, $n_1 = n_2 = n_3 = 20$

Next, we shall look at the Cornish—Fisher expansion of T_W when the data can be divided into two identifiable strata S_1 and S_2 for simplicity (these results can be extended to p strata in general along the same lines), where each stratum is assumed to be normal, and obtain equivalent degrees of freedom to approximate the distribution of T_W, using the Cornish—Fisher expansion of T_W, where the weights $\{w_i\}$ are fixed.

Let n_1 be the number of observations coming from S_1 and n_2 be the number of observations coming from S_2. Further, we assume that

$$\frac{n_i}{n} \to \theta_i \qquad (i=1,2). \tag{3.45}$$

Each observation in S_k is given a weight w_k; $k=1,2$, where $n_1 w_1 + n_2 w_2 = 1$. Now let

$$\bar{Y}_W = w_1 \sum_{Y_j \in S_1} Y_j + w_2 \sum_{Y_j \in S_2} Y_j ;$$

It is clear that under the assumption $\frac{n_i}{n} \to \theta_i$, $w_i \sim O(\frac{1}{n})$. Before we proceed we introduce the following notations.

$$\begin{aligned}
\tau_1^2 &= \mathrm{var}(Y_j - \bar{Y}_W) ; && \text{if } Y_j \in S_1, \\
\tau_2^2 &= \mathrm{var}(Y_j - \bar{Y}_W) ; && \text{if } Y_j \in S_2, \\
\rho_1 &= \mathrm{Corr}(Y_j - \bar{Y}_W, Y_k - \bar{Y}_W); && \text{if } Y_j, Y_k \in S_1,\ j \neq k, \\
\rho_2 &= \mathrm{Corr}(Y_j - \bar{Y}_W, Y_k - \bar{Y}_W); && \text{if } Y_j, Y_k \in S_2,\ j \neq k,
\end{aligned}$$

$$\rho_{12} = \mathrm{Corr}(Y_j - \bar{Y}_W, Y_k - \bar{Y}_W);\ \text{if } Y_j \in S_1 \text{ and } Y_k \in S_2 \text{ or } Y_j \in S_2 \text{ and } Y_k \in S_1.$$

Hence we obtain

$$\begin{aligned}\tau_1^2 &= (1-2w_1)\sigma_1^2 + n_1 w_1^2\sigma_1^2 + n_2 w_2^2\sigma_2^2 \\ &= (1-2w_1)\sigma_1^2 + n\theta_1 w_1^2\sigma_1^2 + n\theta_2 w_2^2\sigma_2^2 \\ &= \sigma_1^2 + O(\tfrac{1}{n}).\end{aligned}$$

and

$$\begin{aligned}\tau_2^2 &= (1-2w_2)\sigma_2^2 + n_1 w_1^2\sigma_1^2 + n_2 w_2^2\sigma_2^2 \\ &= (1-2w_2)\sigma_2^2 + n\theta_1 w_1^2\sigma_1^2 + n\theta_2 w_2^2\sigma_2^2. \\ &= \sigma_2^2 + O(\tfrac{1}{n}).\end{aligned}$$

After some algebra one can show that

$$\rho_1 = \frac{(n\theta_1 w_1^2\sigma_1^2 + n\theta_2 w_2^2\sigma_2^2 - 2w_2\sigma_2^2)}{\sigma_1^2} + O(\frac{1}{n^2}),$$

$$\rho_2 = \frac{(n\theta_1 w_1^2\sigma_1^2 + n\theta_2 w_2^2\sigma_2^2 - 2w_1\sigma_1^2)}{\sigma_2^2} + O(\frac{1}{n^2}),$$

and

$$\rho_{12} = \frac{(n\theta_1 w_1^2\sigma_1^2 + n\theta_2 w_2^2\sigma_2^2 - w_1\sigma_1^2 - w_2\sigma_2^2)}{\sigma_1\sigma_2} + O(\frac{1}{n^2}),$$

Cressie (1982) shows that if $w_i \sim O(\frac{1}{n})$ then $\delta_i \sim O(\frac{1}{n^2})$. Hence we obtain

$$\begin{aligned}\mathrm{var}(S^2_{w,\delta}) &= 2[n\theta_1\delta_1^2\tau_1^4(1 + (n\theta_1 - 1)\rho_1^2) + n\theta_2\delta_2^2\tau_2^4(1 + (n\theta_2 - 1)\rho_2^2) \\ &\qquad + 2n^2\theta_1\theta_2\delta_1\delta_2\tau_1^2\tau_2^2\rho_{12}^2] \\ &= 2n\theta_1\delta_1^2\sigma_1^4 + 2n\theta_2\delta_2^2\sigma_2^4 + O(\frac{1}{n^4}).\end{aligned}$$

Now since we chose $\{\delta_i\}$ such that $E(S^2_{w,\delta}) = \mathrm{var}(\bar{Y}_w)$, we obtain

$$E(S^2_{w,\delta}) = n\theta_1 w_1^2\sigma_1^2 + n\theta_2 w_2^2\sigma_2^2 .$$

Hence the Cornish—Fisher expansion of $S^2_{w,\delta}$ is:

$$CF(S^2_{w,\delta}) = (n\theta_1 w_1^2\sigma_1^2 + n\theta_2 w_2^2\sigma_2^2) + (2n\theta_1\delta_1^2\sigma_1^4 + 2n\theta_2\delta_2^2\sigma_2^4)^{1/2}\eta + O(\frac{1}{n^4})$$

$$= (n\theta_1 w_1^2\sigma_1^2 + n\theta_2 w_2^2\sigma_2^2)\left[1 + \frac{(2n\theta_1\delta_1^2\sigma_1^4 + 2n\theta_2\delta_2^2\sigma_2^4)^{1/2}}{(n\theta_1 w_1^2\sigma_1^2 + n\theta_2 w_2^2\sigma_2^2)}\eta\right] + O(\frac{1}{n^4}), \tag{3.46}$$

where η is a standard normal random variable.

Also we can write the Cornish—Fisher expansion of $\bar{Y}_W$ as

$$\bar{Y}_W = \mu + (n\theta_1 w_1^2\sigma_1^2 + n\theta_2 w_2^2\sigma_2^2)^{1/2}\varsigma + O(\frac{1}{n^2}), \tag{3.47}$$

where ς is a standard normal random variable.

Now since

$$T_W = \frac{\bar{Y}_W - \mu}{S_{w,\delta}},$$

and observing that $\bar{Y}_W$ is uncorrelated with $S^2_{w,\delta}$ for symmetric populations (in particular under normality) we obtain the Cornish—Fisher expansion of T_W :

$$CF(T_W) = (n\theta_1 w_1^2\sigma_1^2 + n\theta_2 w_2^2\sigma_2^2)^{1/2}\varsigma \cdot (n\theta_1 w_1^2\sigma_1^2 + n\theta_2 w_2^2\sigma_2^2)^{-1/2} \cdot \left[1 + \frac{(2n\theta_1\delta_1^2\sigma_1^4 + 2n\theta_2\delta_2^2\sigma_2^4)^{1/2}}{(n\theta_1 w_1^2\sigma_1^2 + n\theta_2 w_2^2\sigma_2^2)}\eta\right]^{-1/2} + O(\frac{1}{n^2})$$

$$= \varsigma \cdot \left[1 - \frac{1}{2}\left(\frac{2}{n}\right)^{1/2}\frac{(\theta_1\delta_1^2\sigma_1^4 + \theta_2\delta_2^2\sigma_2^4)^{1/2}}{(\theta_1 w_1^2\sigma_1^2 + \theta_2 w_2^2\sigma_2^2)}\eta\right] + O(\frac{1}{n^2}). \tag{3.48}$$

Where ς and η are independent standard normal random variables.

The above Cornish—Fisher expansion is valid for normal populations. To obtain an equivalent degrees of freedom for T_W we could compare the Cornish—Fisher expansion of T_W given by (3.48) to the Cornish—Fisher expansion of a random variable which has a Student's t—distribution with f_W degrees of freedom. Suppose T_W follows an approximate t—distribution with f_W degrees of freedom. Then we should be able to write

$$T_W = Z \cdot \left\{1 - \frac{1}{\sqrt{2f_W}} Z^*\right\} + O\left(\frac{1}{f_W^2}\right). \tag{3.49}$$

Comparing (3.48) and (3.49) we obtain

$$f_W = n \frac{(\theta_1 w_1^2 \sigma_1^2 + \theta_2 w_2^2 \sigma_2^2)^2}{(\theta_1 \delta_1^2 \sigma_1^4 + \theta_2 \delta_2^2 \sigma_2^4)}. \tag{3.50}$$

In a more general setting, where we have p different strata such that $\frac{n_i}{n} \to \theta_i$ $(i = 1, 2, \ldots, p)$, along the same lines as above one can show that

$$f_W = n \frac{\left(\sum_{i=1}^{p} \theta_i w_i^2 \sigma_i^2\right)^2}{\left(\sum_{i=1}^{p} \theta_i \delta_i^2 \sigma_i^4\right)}. \tag{3.51}$$

A quick look at of (3.51) shows that if $\sigma_i^2 = \sigma^2$and $w_i = \frac{1}{n}$ then $f = n$ which agrees to the leading order. Also since $w_i \sim O(\frac{1}{n})$ we see from (3.41) that $\delta_i \approx w_i^2$ and therefore to $O(1)$ we have

$$f_W = n \frac{\left(\sum_{i=1}^{p} \theta_i \delta_i \tau_i^2\right)^2}{\left(\sum_{i=1}^{p} \theta_i \delta_i^2 \tau_i^2\right)}. \tag{3.52}$$

We see immediately the equivalence of the equivalent degrees of freedom formulae given by (3.44) and (3.52). Cressie (1982) used a method suggested by Satterthwaite (1946) to obtain the e.d.f. given by (3.44). We used the Cornish–Fisher expansion to obtain the e.d.f. given by (3.52). Hence we can conclude that both these approaches lead to the same approximation of T_W as a t–distribution.

3.5. M–estimate of the Common Mean μ

In this section we shall consider M–estimates of the common mean μ and present some results on asymptotically safe test statistics for use in testing and constructing confidence intervals for the common mean μ. For completeness we shall give a brief introduction to M–estimation of a location parameter. We refer the reader to Huber (1964, 1981) for a detailed discussion on this subject.

Let the sample observations $Y_1, Y_2, \ldots, Y_n$ be generated from a family F_θ; $\theta \in \Theta$. We wish to estimate the unknown parameter θ. If the parametric family is known one usually employs maximum likelihood estimation; i.e., maximize

$$\sum_{i=1}^{n} \log f_\theta(Y_i),$$

with respect to θ. If it is differentiable with respect to θ, this is equivalent to solving

$$\sum_{i=1}^{n} \Psi(Y_i;\hat{\theta}) = 0,$$

where

$$\Psi(y;\theta) = \frac{d}{d\theta}\log f_\theta(y).$$

An estimation procedure which generalizes maximum likelihood estimation is that of M—estimation: the "M" stands for the maximum—likelihood—type estimation. In this procedure we try to solve

$$\sum_{i=1}^{n} \Psi(Y_i;\hat{\theta}) = 0,$$

where Ψ is a pre-chosen function.

In particular when our interest centers around a location estimate then we solve

$$\sum_{i=1}^{n} \Psi(Y_i - \hat{\theta}) = 0. \tag{3.53}$$

The above equation can be equivalently written as

$$\sum_{i=1}^{n} w_i \cdot (Y_i - \hat{\theta}) = 0,$$

where

$$w_i = \frac{\Psi(Y_i - \hat{\theta})}{Y_i - \hat{\theta}}.$$

Hence one could represent $\hat{\theta}$ as a weighted mean, i.e., $\hat{\theta} = \frac{\sum_{i=1}^{n} w_i Y_i}{\sum_{i=1}^{n} w_i}$, with weights depending on the estimator itself. This representation is particularly useful when iterative procedures are to be employed to solve (3.53).

As an example Kafadar (1982) considers a Ψ—function defined by

$$\Psi(u) = \begin{cases} u(1-u^2)^2 & \text{if } |u| \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

and solves

$$\sum_{i=1}^{n} \Psi\left(\frac{Y_i - \hat{\theta}}{cS}\right) = 0, \tag{3.54}$$

using iterative procedures, where S is an estimate of scale and c is a suitably chosen constant. This yields an M—estimator of the location parameter θ, which is commonly referred to as the biweight estimator of θ. The rationale behind using cS in (3.54) is to make the estimator scale invariant.

Mosteller and Tukey (1977) recommend a value of c when S is chosen to be the median absolute deviation so that cS is between 4σ and 6σ if it happens that the observations are normal and identically distributed with scale parameter σ. Kafadar then constructs a "t"—like statistic and shows via Monte Carlo simulations that this statistic is efficient in terms of the expected length of the confidence intervals for samples of moderate sizes.

As we are interested in estimating the common mean μ when the homoskedasticity assumption is relaxed, we propose to use the weighted M—estimator suggested by Cressie (1980b). Henceforth in this section we shall assume that the observations can be divided into p identifiable strata so that equal variation occurs in each stratum; i.e., assume that

$$\frac{Y_{ij}-\mu}{\sigma_i} \sim G; \qquad (i=1,2,\ldots,p;\ j=1,2,\ldots,n_i),$$

where G has mean 0 and variance 1.

3.5.1. Definition Cressie (1982)

The weighted M—estimator $\tilde{Y}_{W,V,\Psi}$ of μ is defined as the solution for $\tilde{Y}$ in

$$\sum_{i=1}^{p} w_i \sum_{j=1}^{n_i} \Psi(v_i(Y_{ij}-\tilde{Y})) = 0,$$

where $\{w_i \geq 0; i=1,2,\ldots,p\}$, $\{v_i \geq 0; i=1,2,\ldots,p\}$ and Ψ is a pre—chosen function.

The weights $\{w_i\}$, and $\{v_i\}$ are referred to as the external and internal weights respectively. The following theorem due to Cressie (1980b), which modifies a proof for the unweighted case given by Huber (1964), gives the asymptotic normality of the weighted M—estimator $\tilde{Y}_{W,V,\Psi}$ of μ. We shall be using this theorem in the subsequent discussions.

<u>3.5.2.</u> <u>Theorem</u> Cressie (1982)

Define $\lambda(\xi) \equiv \sum_{i=1}^{p}\theta_i w_i E(\Psi(v_i(Y_{ij}-\xi)))$,

where $\frac{n_i}{n} \rightarrow \theta_i$ as $n \rightarrow \infty$, $n = \sum_{i=1}^{p} n_i$ and for h any measurable function

$E(h(Y_{ij})) \equiv \int h(x)dG((x-\mu)/\sigma_i)$. Assume that (*Assumptions* (*a*))

(i) Ψ is a monotone nondecreasing function which is strictly positive (negative) for large positive (negative) values of its argument;

(ii) there is a c such that $\lambda(c)=0$;

(iii) $\lambda(\xi)$ is differentiable at $\xi = c$, and $\lambda'(c) < 0$;

(iv) $\int \Psi^2(v_j(x-\xi))dG((x-\mu)/\sigma_i)$ is finite and continous at $\xi = c$; $j=1,2,\ldots,p$.

or assume that (*Assumptions* (*b*))

(i) $\tilde{Y}_{W,V,\Psi} \rightarrow c$ in probability;

(ii) Ψ is continuous and has a uniformly continuous derivative Ψ';

(iii) $\int \Psi^2(v_i(x-\mu))dG((x-\mu)/\sigma_i) < \infty \quad (i=1,2,\ldots,p);$

(iv) $0 < \int \Psi'(v_i(x-\mu))dG((x-\mu)/\sigma_i) < \infty \quad (i=1,2,\ldots,p).$

Then

$$n^{1/2}(\tilde{Y}_{W,V,\Psi}-c) \xrightarrow{d} N\left(0, \frac{\sum_{i=1}^{p} w_i^2\theta_i E[\Psi^2(v_i(Y_{ij}-c))]}{\{\sum_{i=1}^{p} w_i\theta_i v_i E[\Psi'(v_i(Y_{ij}-\mu))]\}^2}\right).$$

□

As a special case let us further assume that G is symmetric about 0. In this situation one can replace c by μ in the above theorem. Let us consider the external and internal weights given by $w_i=1$ and $v_i=\frac{1}{\hat{\sigma}}$ for $i=1,2,\ldots,p$, where $\hat{\sigma}$ is given by

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{p}\sum_{j=1}^{n_i}(Y_{ij}-\bar{Y})^2}{(n-1)}.$$

Before we proceed let us prove the following lemma.

3.5.3. Lemma

Assume $n_i/n \to \theta_i$, as $n\to\infty$, where $\sum_{i=1}^{p}\theta_i=1$. Then

$$\hat{\sigma}^2 \to \sum_{i=1}^{p}\theta_i\sigma_i^2 \equiv \sigma^2, \text{ in probability, as } n\to\infty.$$

Proof:

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{p}\sum_{j=1}^{n_i}(Y_{ij}-\bar{Y})^2}{(n-1)}$$

$$= \sum_{i=1}^{p}\frac{(n_i-1)}{(n-1)}\sum_{j=1}^{n_i}\frac{(Y_{ij}-\bar{Y})^2}{(n_i-1)}$$

$$\to \sum_{i=1}^{p}\theta_i\sigma_i^2, \text{ in probability.}$$

Q.E.D

Hence we have the following theorem.

3.5.4. Theorem

For the special case of the external and internal weights given by $w_i = 1$ and $v_i = \frac{1}{\sigma}$; $i = 1,2,\ldots,p$, let the resulting weighted M—estimator be denoted by $\tilde{Y}_\Psi$, then

$$n^{1/2}(\tilde{Y}_\Psi - \mu) \xrightarrow{d} N\left(0, \frac{\sum_{i=1}^{p} \theta_i E[\Psi^2((Y_{ij} - \mu)/\sigma)]}{\{\sum_{i=1}^{p} \frac{\theta_i}{\sigma} E[\Psi'((Y_{ij} - \mu)/\sigma)]\}^2}\right),$$

where σ^2 is defined in Lemma 3.5.3.

Proof

The proof immediately follows from Theorem 3.5.2, Lemma 3.5.3 and Slutsky's theorem.

Q.E.D.

For finite sample considerations let us consider the asymptotic equivalence form of the Studentized M—estimator defined by

$$T_\Psi = \frac{n^{1/2}(\tilde{Y}_\Psi - \mu)}{D_\Psi}, \tag{3.55}$$

where

$$D_\Psi^2 = \frac{\frac{1}{n} \sum_{i=1}^{p} \sum_{j=1}^{n_i} \Psi^2((Y_{ij} - \tilde{Y}_\Psi)/\hat{\sigma})}{\left(\frac{1}{n} \sum_{i=1}^{p} \sum_{j=1}^{n_i} \frac{1}{\hat{\sigma}} \Psi'((Y_{ij} - \tilde{Y}_\Psi)/\hat{\sigma})\right)^2}.$$

3.5.5. Theorem

T_Ψ given by (3.55) is asymptotically safe; i.e.,

$$D_\Psi^2 \rightarrow \frac{\sum_{i=1}^{p} \theta_i E[\Psi^2((Y_{ij}-\mu)/\sigma)]}{\{\sum_{i=1}^{p} \frac{\theta_i}{\sigma} E[\Psi'((Y_{ij}-\mu)/\sigma)]\}^2} \quad \text{in pr.,}$$

and

$$T_\Psi \rightarrow N(0,1).$$

Proof

First consider the numerator of D_Ψ^2.

$$\frac{1}{n} \sum_{i=1}^{p} \sum_{j=1}^{n_i} \Psi^2((Y_{ij}-\tilde{Y}_\Psi)/\hat{\sigma}) = \sum_{i=1}^{p} \frac{n_i}{n} \sum_{j=1}^{n_i} \frac{\Psi^2((Y_{ij}-\tilde{Y}_\Psi)/\hat{\sigma})}{n_i}.$$

Now by Lemma (3.5.3) $\hat{\sigma} \rightarrow \sigma$ in probability and also we assumed that $\tilde{Y}_\Psi \rightarrow \mu$ in probability. Therefore by the Weak Law of Large Numbers we get

$$\sum_{j=1}^{n_i} \frac{\Psi^2((Y_{ij}-\tilde{Y}_\Psi)/\hat{\sigma})}{n_i} \rightarrow E[\Psi^2((Y_{ij}-\mu)/\sigma)].$$

Observing that $\frac{n_i}{n} \rightarrow \theta_i$ we obtain

$$\frac{1}{n} \sum_{i=1}^{p} \sum_{j=1}^{n_i} \Psi^2((Y_{ij}-\tilde{Y}_\Psi)/\hat{\sigma}) \rightarrow \sum_{i=1}^{p} \theta_i E[\Psi^2((Y_{ij}-\mu)/\sigma)] \text{ in pr.}$$

Again, considering the denominator of D^2_{Ψ} and by a similar reasoning we get

$$\frac{1}{n}\sum_{i=1}^{p}\sum_{j=1}^{n_i}\frac{1}{\hat{\sigma}}\Psi'((Y_{ij}-\tilde{Y}_{\Psi})/\hat{\sigma}) \rightarrow \sum_{i=1}^{p}\frac{\theta_i}{\sigma}E[\Psi'((Y_{ij}-\mu)/\sigma)], \text{ in pr.}$$

$$D^2 \rightarrow \frac{\sum_{i=1}^{p}\theta_i E[\Psi^2((Y_{ij}-\mu)/\sigma)]}{\{\sum_{i=1}^{p}\frac{\theta_i}{\sigma}E[\Psi'((Y_{ij}-\mu)/\sigma)]\}^2} \quad \text{in pr.}$$

Therefore by Slutsky's theorem we get $T_{\Psi} \xrightarrow{d} N(0,1)$.

Q.E.D

For any set of fixed internal and external weights we can define $T_{W,V,\Psi}$ analogously, i.e., let

$$T_{W,V,\Psi} = \frac{n^{1/2}(\tilde{Y}_{W,V,\Psi}-\mu)}{D_{W,V,\Psi}}, \tag{3.56}$$

where

$$D^2_{W,V,\Psi} = \frac{\frac{1}{n}\sum_{i=1}^{p}\sum_{j=1}^{n_i} w_i^2\Psi^2(v_i(Y_{ij}-\tilde{Y}_{W,V,\Psi}))}{\{\frac{1}{n}\sum_{i=1}^{p}\sum_{j=1}^{n_i} w_i v_i \Psi'(v_i(Y_{ij}-\tilde{Y}_{W,V,\Psi}))\}^2}.$$

3.5.6. Theorem

$T_{W,V,\Psi}$ given by (3.56) is asymptotically safe and $T_{W,V,\Psi} \xrightarrow{d} N(0,1)$.

Proof

Since the weights $\{w_i\}$ and $\{v_i\}$ are fixed, the proof immediately follows by the Weak Law of Large Numbers and Slutsky's theorem.

Q.E.D

Let us summarize the results obtained in this chapter. We established two new theorems concerning combining two unbiased estimators of a common mean and obtained an upper bound for the inefficiency of such an estimator. Combining more than two unbiased estimators in order to obtain a better unbiased estimator is under investigation. We also discussed the notion of *safe* T—statistics introduced by Cressie (1982) and used his ideas to construct *asymptotically safe* test statistics for inferential problems concerning a common mean μ, using weighted M—estimates.

4. LINEAR MODEL IN THE PRESENCE OF HETEROSKEDASTICITY

4.1. Introduction

In the process of learning and understanding nature, modeling is considered to be one of the starting points. We might like our model to fit reality exactly, but we all know with the limited knowledge we have, that this is impossible in practical situations. One branch of statistics that has proved to be useful in practical modeling situations is linear models.

In the theory of linear models it is well known that heteroskedasticity in error terms leads to consistent but often inefficient parameter estimates and inconsistent covariance matrix estimates. In such situations faulty inference may be drawn.

If we knew the structure of the heteroskedasticity we could overcome the difficulty above by performing a suitable transformation on the data. But this is not a common situation that we come across in practice. White (1980a) presents a covariance matrix estimator that is consistent in the presence of heteroskedasticity, but it does not rely on a specific formal model of the structure of heteroskedasticity. This enables us to make valid asymptotic inference even when the linear transformation on the data is either unknown or incorrect.

This chapter has as its basis, White's results. In Section 4.2 we shall consider general results in the fixed regressors situation, although White also considers stochastic regressors. These will be applied to the 1—sample problem in Section 4.3 and the 2—sample problem in Section 4.4. Finally Section 4.5 will be devoted to the application of White's results to the simple linear regression problem with and without intercept.

4.2. Heteroskedasticity—Consistent Covariance Matrix Estimator and Related Results

Before we present White's (1980a) results let us introduce the necessary notation. Assume the following structure for the model.

4.2.1. Assumption

Let

$$Y_i = \underline{X}_i' \beta_o + \epsilon_i \qquad (i = 1,2,\ldots,n), \tag{4.1}$$

where $\{\epsilon_i : i = 1,2,\ldots,n\}$ is a sequence of independent but not necessarily identically distributed (i.n.i.d.) random errors such that $E(\epsilon_i) = 0$, $E(\epsilon_i^2) = \sigma_i^2$; $i = 1,2,\ldots,n$, $\underline{X}_i$ is a kx1 vector of deterministic components (fixed regressors), and β_o is a finite kx1 parameter vector to be estimated.

By assuming the ϵ_i's are i.n.i.d. the case of heteroskedastic errors is automatically covered.

4.2.2. Assumption

There exist positive constants δ and Δ such that

(i) $E(|\epsilon_i^2|^{1+\delta}) < \Delta \qquad (i=1,2,\ldots,n),$

(ii) $|X_{ij}| < \Delta \qquad (i=1,2,\ldots,n;\ j=1,2,\ldots,k;\ n=1,2,\ldots).$

and

(iii) $\bar{M}_n = \frac{1}{n}\sum_{i=1}^{n} \underline{X}_i \underline{X}_i'$ is nonsingular for (all) n sufficiently large and $\det \bar{M}_n > \delta > 0$.

Let

$$\hat{\underline{\beta}}_n = (X'X)^{-1}X'\underline{Y}, \tag{4.2}$$

be the ordinary least squares (o.l.s.) estimator of $\underline{\beta}_0$, where

$$X = \begin{bmatrix} \underline{X}_1' \\ \underline{X}_2' \\ \vdots \\ \underline{X}_n' \end{bmatrix},$$

is the nxk design matrix, and

$$\underline{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix},$$

is the nx1 vector of observations.

Then

$$\begin{aligned} \operatorname{var}(\hat{\underline{\beta}}_n) &= (X'X)^{-1}X'\Omega X(X'X)^{-1} \\ &= \frac{1}{n}\left(\frac{X'X}{n}\right)^{-1}\left(\frac{X'\Omega X}{n}\right)^{-1}\left(\frac{X'X}{n}\right)^{-1}. \end{aligned} \tag{4.3}$$

where $\Omega = \operatorname{diag}(\sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2)$ and $\sigma_i^2 = E(\epsilon_i^2)$ $(i=1,2,\ldots,n)$.

4.2.3 Lemma (White, 1980a)

Under the Assumptions 4.2.1 and 4.2.2, $\hat{\underline{\beta}}_n$ exists for sufficiently large n and is a strongly consistent estimator of $\underline{\beta}_o$ (i.e., $\hat{\underline{\beta}}_n \xrightarrow{a.s} \underline{\beta}_o$).

Proof:

White's proof for stochastic regressors, is modified here in the special case of fixed regressors. We observe that $\bar{M}_n = \frac{1}{n}(X'X)$, and since $\bar{M}_n$ is nonsingular under Assumption 4.2.2 (iii) we can write

$$\begin{aligned}
\hat{\underline{\beta}}_n &= (X'X)^{-1}X'\underline{Y} \\
&= (X'X/n)^{-1}(X'(X\underline{\beta}_o + \underline{\epsilon})/n) \\
&= \underline{\beta}_o + (X'X/n)^{-1}(X'\underline{\epsilon}/n) \\
&= \underline{\beta}_o + \bar{M}_n^{-1}(X'\underline{\epsilon}/n) \qquad (4.4)
\end{aligned}$$

where

$$\underline{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}.$$

From (4.4) we immediately see that since $\bar{M}_n$ is nonsingular $\hat{\underline{\beta}}_n$ exists. Now since the $\underline{X}_i \epsilon_i$'s are independent, by the Markov strong law of large numbers (see Chung, 1973, pp. 125-126) we obtain

$$|(X'\underline{\epsilon}/n) - n^{-1}\sum_{i=1}^{n} E(\underline{X}_i \epsilon_i)| \xrightarrow{a.s} 0,$$

i.e.,

$$|(X'\underline{\epsilon}/n) - n^{-1}\sum_{i=1}^{n} \underline{X}_i E(\epsilon_i)| \xrightarrow{a.s} 0.$$

But $E(\epsilon_i) = 0$ under Assumption 4.2.1. Thus

$$X'\underline{\epsilon}/n \xrightarrow{a.s} 0.$$

Now we will show that $\bar{M}_n^{-1}$ has uniformly bounded elements under the Assumption 4.2.2 (ii) and (iii). Recall that

$$\bar{M}_n = \frac{1}{n} X'X.$$

Therefore the ijth element of $\bar{M}_n$ is $\bar{m}_{n,ij} = \frac{1}{n}\sum_{k=1}^{n} X_{ki}X_{kj}$, and $|\bar{m}_{n,ij}| \leq \Delta^2$ under the Assumption 4.2.2 (ii).

Now if we denote the cofactor of $\bar{m}_{n,ij}$ by $\bar{M}_{n,ij}$, it is immediate that under Assumption 4.2.2 (ii), $|\bar{M}_{n,ij}| \leq \Delta^{2(k-1)}$. Also under the Assumption 4.2.2 (iii), $\det \bar{M}_n > \delta > 0$. Therefore the ijth element of $\bar{M}_n^{-1} \leq \frac{\Delta^{2(k-1)}}{\delta}$ and hence $\bar{M}_n^{-1}$ has uniformly bounded elements. Thus

$$\bar{M}_n^{-1} X' \underline{\epsilon}/n \overset{a.s}{\rightarrow} 0,$$

and hence

$$\hat{\underline{\beta}}_n \overset{a.s}{\rightarrow} \underline{\beta}_o.$$

Q.E.D.

In order to obtain an asymptotic normality result for $\hat{\underline{\beta}}_n$, we introduce the following assumption.

<u>4.2.4. Assumption</u>

There exists a positive finite constant δ such that

$$\bar{V}_n = \frac{1}{n} \sum_{i=1}^{n} (\underline{X}_i \underline{X}_i') E(\epsilon_i^2) = \frac{X'\Omega X}{n},$$

is nonsingular for sufficiently large n and $\det \bar{V}_n > \delta > 0$, where $\Omega = \text{diag}(\sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2)$ and $\sigma_i^2 = E(\epsilon_i^2)$; $i = 1, 2, \ldots, n$.

The above assumption and Assumption 4.2.2 (ii) ensures the uniform boundedness of the elements of $\bar{V}_n^{-1}$.

4.2.5. Lemma (White, 1980a)

Under Assumptions 4.2.1, 4.2.2, and 4.2.4,

$$n^{1/2}\bar{V}_n^{-1/2}\bar{M}_n(\hat{\underline{\beta}}_n - \underline{\beta}_o) \xrightarrow{d} N(\underline{0}, I_k),$$

where $\bar{V}_n^{-1/2}$ is a symmetric positive definite matrix such that $(\bar{V}_n^{-1/2})^2 = \bar{V}_n^{-1}$, and I_k is the kxk identity matrix.

□

We omit the proof of this lemma and refer the reader to White (1980a). Notice this result is slightly more general than the usual asymptotic normality result since the covariance matrix

$$\frac{\bar{M}_n^{-1}\bar{V}_n\bar{M}_n^{-1}}{n} = \frac{1}{n}\left(\frac{X'X}{n}\right)^{-1}\left(\frac{X'\Omega X}{n}\right)^{-1}\left(\frac{X'X}{n}\right)^{-1},$$

need not necessarily converge to any particular matrix.

We could now apply the lemma above to form a test statistic to test the linear hypotheses of the form

$$H_o: R\underline{\beta}_o = \underline{r} \qquad \text{vs} \qquad H_1: R\underline{\beta}_o \neq \underline{r},$$

where R is a finite qxk matrix of full row rank and $\underline{r}$ is a qx1 vector. It can be shown (White, 1980a) that under H_o

$$n(R\hat{\underline{\beta}}_n - \underline{r})'[R\bar{M}_n^{-1}\bar{V}_n\bar{M}_n^{-1}R']^{-1}(R\hat{\underline{\beta}}_n - \underline{r}) \xrightarrow{d} \chi_q^2.$$

Here we should note that the above statistic is not computable since $\bar{V}_n$ depends on unknown parameters; if we could replace it with a consistent estimator then we could perform the usual asymptotic tests.

In the situation we are interested in $\bar{V}_n = \frac{1}{n}\sum_{i=1}^{n} E(\underline{X}_i \epsilon_i^2 \underline{X}_i') = \frac{X'\Omega X}{n}$, where $\Omega = \mathrm{diag}(\sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2)$ and $\sigma_i^2 = E(\epsilon_i^2)$ $(i = 1, 2, \ldots, n)$. We might think that this is an impossible task since we have only n observations. Actually we do not have to estimate n different variances separately, but rather an average of expectations. Under the conditions set out earlier, a natural candidate to estimate this average of expectations is $\frac{1}{n}\sum_{i=1}^{n}(\underline{X}_i \epsilon_i^2 \underline{X}_i')$. Again we see that this is also impossible due to the fact that ϵ_i's are unknown. However, the ϵ_i's can be estimated by the ordinary least squares residuals $\hat{\epsilon}_i = Y_i - \underline{X}_i'\hat{\underline{\beta}}_n$. Therefore an estimator for the $\mathrm{var}(\hat{\underline{\beta}}_n)$ given by (4.3) can be obtained.

$$\hat{\mathrm{var}}(\hat{\underline{\beta}}_n) = \frac{\bar{M}_n^{-1}\hat{V}_{n,w}\bar{M}_n^{-1}}{n} \equiv \tilde{V}_{n,w} \text{ (say)}, \tag{4.5}$$

where $\hat{V}_{n,w} = \frac{1}{n}\sum_{i=1}^{n}(\underline{X}_i \hat{\epsilon}_i^2 \underline{X}_i') = \frac{X'\hat{\Omega}X}{n}$ and $\hat{\Omega} = \mathrm{diag}(\hat{\epsilon}_1^2, \hat{\epsilon}_2^2, \ldots, \hat{\epsilon}_n^2)$.
The next theorem shows that $\hat{V}_{n,w}$ is a consistent estimator of $\bar{V}_n$.

<u>4.2.6.</u> <u>Theorem</u> (White, 1980a)

Under the Assumptions 4.2.1, 4.2.2, and 4.2.4,

(i) $|\hat{V}_{n,w} - \bar{V}_n| \xrightarrow{a.s} 0$,

(ii) $n^{1/2}\hat{V}_{n,w}^{-1/2}\bar{M}_n(\hat{\underline{\beta}}_n - \underline{\beta}_o) \xrightarrow{d} N(0, I_k)$,

and

(iii) $n(R\hat{\underline{\beta}}_n - \underline{r})'[R\bar{M}_n^{-1}\hat{V}_{n,w}\bar{M}_n^{-1}R']^{-1}(R\hat{\underline{\beta}}_n - \underline{r}) \xrightarrow{d} \chi_q^2$ under H_o. □

Again we refer the reader to White (1980a) for a proof of this theorem. Before we proceed further let us investigate the bias of $\hat{V}_n$. The following notation will be introduced for convenience.

Let

$$H = X(X'X)^{-1}X',$$

$$\begin{aligned} P &= I - X(X'X)^{-1}X' \\ &= I - H. \end{aligned}$$

Then the o.l.s residual vector $\underline{\hat{\epsilon}}_n = (\hat{\epsilon}_1, \hat{\epsilon}_2, \ldots, \hat{\epsilon}_n)'$ can be written as

$$\begin{aligned} \underline{\hat{\epsilon}}_n &= \underline{Y} - X\underline{\hat{\beta}}_n \\ &= \underline{Y} - X(X'X)^{-1}X'\underline{Y} \\ &= (I - X(X'X)^{-1}X')\underline{Y} \\ &= P\underline{Y}. \end{aligned}$$

Therefore

$$\begin{aligned} \mathrm{var}(\underline{\hat{\epsilon}}_n) &= \mathrm{var}(P\underline{Y}) \\ &= P\Omega P' \\ &= P\Omega P, \end{aligned}$$

since P is symmetric.

Now noticing that $E(\underline{\hat{\epsilon}}_n) = \underline{0}$, we obtain

$$\begin{aligned} E(\hat{\epsilon}_i^2) &= \underline{p}_i'\Omega\underline{p}_i \\ &= (\underline{e}_i' - \underline{h}_i')\Omega(\underline{e}_i - \underline{h}_i) \\ &= \sigma_i^2 - 2\sigma_i^2\underline{h}_i'\underline{e}_i + \underline{h}_i'\Omega\underline{h}_i \\ &= \sigma_i^2 - 2\sigma_i^2 h_{ii} + \underline{h}_i'\Omega\underline{h}_i \\ &= \sigma_i^2 - 2\sigma_i^2\underline{h}_i'\underline{h}_i + \underline{h}_i'\Omega\underline{h}_i, \end{aligned} \tag{4.6}$$

where $\underline{e}_i$ and $\underline{h}_i$ represents the i^{th} columns of I_{nxn} and H respectively and h_{ii} is the i^{th} element of $\underline{h}_i$. Therefore

$$E(\hat{V}_{n,w} - \bar{V}_n) = \frac{X'BX}{n}, \tag{4.7}$$

where $B = \mathrm{diag}(\underline{h}_i'(\Omega - 2\sigma_i^2 I)\underline{h}_i)$.

Now from (4.7) it is immediate that if $\dfrac{\max(\sigma_i^2;\ i=1,2,\ldots,n)}{\min(\sigma_i^2;\ i=1,2,\ldots,n)} < 2$ then all the diagonal elements of B are negative; in particular when errors are homoskedastic, $\hat{V}_{n,w}$ has a negative bias.

Let us look at the bias of $\hat{V}_{n,w}$ in the homoskedastic error variance situation. In this situation all σ_i^2 's are equal (say σ^2). Therefore we see from (4.6) that

$$E(\hat{\epsilon}_i^2) = (1 - h_{ii})\sigma^2.$$

Hence it is obvious that $\hat{\epsilon}_i'^2 = \hat{\epsilon}_i^2/(1 - h_{ii})$ is an unbiased estimator of the common error variance σ^2.

Thus MacKinnon and White (1985) suggested the use of $\hat{\epsilon}_i'$'s instead of $\hat{\epsilon}_i$'s. Weber (1986) shows that White's covariance matrix estimator given by (4.5) and the weighted jackknife covariance matrix estimator proposed by Hinkley (1977) differ only by a scaling factor. This connection was first conjectured by Cressie (1982). Further, Weber also shows that White's covariance matrix estimator and Hinkley's weighted jackknife covariance matrix estimator for $\mathrm{var}(\hat{\underline{\beta}}_n)$ underestimates the variance of components of $\hat{\underline{\beta}}_n$ even if used in a homoskedastic error variance situation. Following a suggestion made by Wu(1986) to reduce the bias in the weighted jackknife estimator and having observed the connection between White's covariance matrix estimator and Hinkley's weighted jackknife covariance matrix estimator, Weber also makes the same suggestion as MacKinnon and White (1985). Hence we obtain another estimator for the var($\hat{\beta}_n$), namely

$$\tilde{V}_{n,mw} = \frac{\bar{M}_n^{-1}\hat{V}_{n,mw}\bar{M}_n^{-1}}{n}, \tag{4.8}$$

where $\hat{V}_{n,mw} = \frac{1}{n}\sum_{i=1}^{n}(\underline{X}_i\hat{\epsilon}_i'^2\underline{X}_i') = \frac{X'\hat{\Omega}'X}{n}$, $\hat{\Omega}' = \text{diag}(\hat{\epsilon}_1'^2, \hat{\epsilon}_2'^2, \ldots, \hat{\epsilon}_n'^2)$ and $\hat{\epsilon}_i'^2 = \hat{\epsilon}_i^2/(1 - h_{ii})$; $i = 1, 2, \ldots, n$.

Since under the Assumption 4.2.2, $\max_{i \le n} h_{ii}$ converges to 0 as $n \to \infty$, the above adjustment does not effect the asymptotic results given by White (1980a). Hence, we have the following theorem.

<u>4.2.7.</u> <u>Theorem</u> (MacKinnon and White, 1985)

Under the Assumptions 4.2.1, 4.2.2, and 4.2.4,

(i) $|\hat{V}_{n,mw} - \bar{V}_n| \xrightarrow{a.s} 0$,

(ii) $n^{1/2}\hat{V}_{n,mw}^{-1/2}\bar{M}_n(\hat{\beta}_n - \beta_o) \xrightarrow{d} N(0, I_k)$,

and

(iii) $n(R\hat{\beta}_n - \underline{r})'[R\bar{M}_n^{-1}\hat{V}_{n,mw}\bar{M}_n^{-1}R']^{-1}(R\hat{\beta}_n - \underline{r}) \xrightarrow{d} \chi_q^2$ under H_o. □

In the following sections, we shall apply the results of Theorem 4.2.7 to the one—sample, two—sample, and simple linear regression problems. Before we proceed we discuss the following two theorems which are concerned with the unbiased estimation of any linear combination of $\{\sigma_i^2 : i = 1, 2, \ldots, n\}$.

4.2.8. Theorem (Rao, 1970)

Let

$$\underline{h}_i = \begin{bmatrix} h_{1i} \\ h_{2i} \\ \vdots \\ h_{ni} \end{bmatrix}_{n \times 1},$$

$$\hat{\underline{\epsilon}}^2 = \begin{bmatrix} \hat{\epsilon}_1^2 \\ \hat{\epsilon}_2^2 \\ \vdots \\ \hat{\epsilon}_n^2 \end{bmatrix}_{n \times 1},$$

$$R = \begin{bmatrix} (1-h_{11})^2 & h_{12}^2 & \cdots & h_{1n}^2 \\ h_{21}^2 & (1-h_{22})^2 & \cdots & h_{2n}^2 \\ \vdots & \vdots & \ddots & \vdots \\ h_{n1}^2 & h_{n2}^2 & \cdots & (1-h_{nn})^2 \end{bmatrix},$$

and

$$\underline{\sigma}^2 = \begin{bmatrix} \sigma_1^2 \\ \sigma_2^2 \\ \vdots \\ \sigma_n^2 \end{bmatrix}_{n \times 1}.$$

If R is invertible, then $\underline{a}'R^{-1}\hat{\underline{\epsilon}}^2$ is an unbiased estimator of $\underline{a}'\underline{\sigma}^2 = \sum_{i=1}^{n} a_i \sigma_i^2$, where

$$\underline{a} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}_{n \times 1}.$$

Proof

Since the H matrix is idempotent it is immediate that $\underline{h}_i' \underline{h}_i = h_{ii}$. Therefore we obtain from (4.6)

$$\begin{aligned} E(\hat{\epsilon}_i^2) &= (1 - 2h_{ii})\sigma_i^2 + \sum_{j=1}^{n} h_{ji}^2 \sigma_j^2 \\ &= (1 - h_{ii})^2 \sigma_i^2 + \sum_{j \neq i}^{n} h_{ji}^2 \sigma_j^2. \end{aligned}$$

Thus we can write

$$E(\hat{\underline{\epsilon}}^2) = R\underline{\sigma}^2.$$

Therefore, when the matrix R is invertible, we obtain

$$E(R^{-1}\hat{\underline{\epsilon}}^2) = \underline{\sigma}^2,$$

and hence

$$E(\underline{\alpha}' R^{-1}\hat{\underline{\epsilon}}^2) = \underline{\alpha}' \underline{\sigma}^2.$$

Q.E.D.

This estimator of $\underline{\alpha}'\underline{\sigma}^2$ is often referred to as the minimum norm quadratic unbiased estimator (MINQUE). We notice here that the elements of the R matrix are obtained simply by squaring the elements of the P matrix defined earlier. Rao (1970) provides sufficient conditions for the existence of R^{-1} and hence for the existence of such estimators. Chew (1970) proposes a class of MINQUE type estimators of $\underline{\alpha}'\underline{\sigma}^2$. The following theorem establishes Chew's result.

4.2.9. Theorem (Chew, 1970)

Let W be an arbitrary diagonal weight matrix,

$$\hat{\underline{\beta}}_W = (X'WX)^{-1}X'W\underline{Y} ,$$

be the corresponding weighted least squares estimator of $\underline{\beta}_0$,

$$\begin{aligned}\hat{\underline{\epsilon}}_W &= \underline{Y} - X\hat{\underline{\beta}}_W \\ &= (I - X(X'WX)^{-1}X'W)\underline{Y} \\ &= K\underline{Y} ,\end{aligned}$$

where $K=(I - X(X'WX)^{-1}X'W)$, be the weighted least squares residual vector, and

$$L=(l_{ij})=(k_{ij}^2);$$

i.e., L is obtained by squaring the elements of K. Then if L^{-1} exists $\underline{\alpha}'L^{-1}\hat{\underline{\epsilon}}_W^2$ is an unbiased estimator of $\underline{\alpha}'\underline{\sigma}^2$, where $\hat{\underline{\epsilon}}_W^2$ is the vector obtained by squaring the elements of the vector $\hat{\underline{\epsilon}}_W$.

Proof:

This follows immediately by a similar argument to that given in the proof of Theorem 4.2.8.

Q.E.D.

When $W=I$, we notice that Chew's and Rao's estimators of $\underline{\alpha}'\underline{\sigma}^2$ are the same. We shall refer to Chew's estimator as the weighted MINQUE estimator. As an application of Theorem 4.2.9 let us consider the problem of constructing a *safe* T—statistc which we already discussed in Section 3.4. There we considered the independent observations $Y_1, Y_2, \ldots, Y_n$ with common mean μ and variances (possibly different) σ_i^2; $i=1,2,\ldots,n$, respectively. That is, write

$$Y_i = \mu + \epsilon_i \qquad (i=1,2,\ldots,n),$$

where $E(\epsilon_i)=0$ and $E(\epsilon_i^2)=\sigma_i^2$; $i=1,2,\ldots,n$. Let W be a diagonal matrix with w_i as the i^{th} diagonal element, where $w_i \geq 0$; $i=1,2,\ldots,n$, and $\sum_{i=1}^{n} w_i = 1$. Then the weighted least squares estimator of μ is

$$\bar{Y}_W = \sum_{i=1}^{n} w_i Y_i,$$

with variance

$$\text{var}(\bar{Y}_W) = \sum_{i=1}^{n} w_i^2 \sigma_i^2.$$

As we discussed in Section 3.4, to obtain a *safe* test statistic, Cressie (1982) obtained an unbiased estimator of $\text{var}(\bar{Y}_W)$ which we denoted by $S^2_{W,\delta}$. We shall show that the variance estimator obtained by Cressie is exactly the same estimator that we obtain using Theorem 4.2.9. In the notation of Theorem 4.2.9,

$$K = \begin{bmatrix} 1-w_1 & -w_2 & \cdots & -w_n \\ -w_1 & 1-w_2 & \cdots & -w_n \\ \vdots & \vdots & \ddots & \vdots \\ -w_1 & -w_2 & \cdots & 1-w_n \end{bmatrix},$$

and

$$L = \begin{bmatrix} (1-w_1)^2 & w_2^2 & \cdots & w_n^2 \\ w_1^2 & (1-w_2)^2 & \cdots & w_n^2 \\ \vdots & \vdots & \ddots & \vdots \\ w_1^2 & w_2^2 & \cdots & (1-w_n)^2 \end{bmatrix}.$$

Then we can write

$$L = A + \underline{U}\underline{V}',$$

where

$$A = \begin{bmatrix} 1-2w_1 & 0 & \cdots & 0 \\ 0 & 1-2w_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1-2w_n \end{bmatrix},$$

$$\underline{U} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix},$$

and

$$\underline{V} = \begin{bmatrix} w_1^2 \\ w_2^2 \\ \vdots \\ w_n^2 \end{bmatrix}.$$

Therefore

$$L^{-1} = A^{-1} - \frac{(A^{-1}\underline{U})(\underline{V}'A^{-1})}{1 + \underline{V}'A^{-1}\underline{U}},$$

see e.g., Rao (1973, p. 33). Using this result and noticing that

$$\hat{\underline{\epsilon}}_w = \begin{bmatrix} Y_1 - \bar{Y}_w \\ Y_2 - \bar{Y}_w \\ \vdots \\ Y_n - \bar{Y}_w \end{bmatrix},$$

we easily obtain the weighted MINQU estimator of $\mathrm{var}(\bar{Y}_w)$, namely,

$$\hat{\mathrm{var}}(\bar{Y}_w) = \left[1 + \sum_{j=1}^{n} \frac{w_j^2}{(1-2w_j)}\right]^{-1} \sum_{i=1}^{n} \frac{w_i^2}{(1-2w_i)}(Y_i - \bar{Y}_w),$$

which we immediately recognize as the estimator $S^2_{w,\delta}$ proposed by Cressie (1982).

4.3. One—Sample Problem

Let us state the problem briefly. Let the sample observations $Y_1, Y_2, \ldots, Y_n$ be independent such that $(Y_i - \mu)/\sigma_i \sim G$ where G is either the standard normal cumulative distribution function (c.d.f.) or some other c.d.f. with mean 0 and variance 1. We are basically interested in making inference about the common mean μ when we have no knowledge of σ_i's. Now we could write

$$Y_i = \mu + \epsilon_i,$$

where $E(\epsilon_i) = 0$, and $E(\epsilon_i^2) = \sigma_i^2 \quad (i = 1, 2, \ldots, n)$.

It is a simple matter to see that this is a special case of the very general situation we discussed in the previous section, and we can easily identify in the context of the general situation that

$$\underline{\beta}_o = \mu,$$

$$X = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}_{n \times 1},$$

$$X'X = n,$$

$$\bar{M}_n = 1,$$

and

$$\bar{V}_n = \frac{\sum_{i=1}^{n} \sigma_i^2}{n}.$$

It can be easily seen that the o.l.s of μ is given by

$$\hat{\mu}_n = \bar{Y}.$$

Thus we obtain

$$\hat{V}_{n,w} = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \bar{Y})^2.$$

Hence by applying Theorem 4.2.6 (ii) we obtain

$$T_{1,n}^{*} \equiv \frac{n^{1/2}\,(\bar{Y} - \mu)}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_i - \bar{Y})^2}} \xrightarrow{d} N(0,1).$$

We immediately notice that the statistic given by $T_{1,n}^{*}$ is not *safe* in finite samples but *asymptotically safe*. Since with $\hat{V}_{n,w}$ the finite sample safeness property does not hold, it is interesting to look at $\hat{V}_{n,mw}$ and use Theorem 4.2.7 (ii) to obtain another statistic. First we observe in this situation that

$$\begin{aligned} h_{ii} &= \underline{X}_i'(X'X)^{-1}\underline{X}_i \\ &= 1 \cdot n^{-1} \cdot 1 \\ &= \frac{1}{n}, \end{aligned}$$

and

$$(1 - h_{ii}) = \frac{(n-1)}{n}.$$

Thus

$$\hat{V}_{n,mw} = \frac{1}{(n-1)}\sum_{i=1}^{n}(Y_i - \bar{Y})^2.$$

Therefore applying Theorem 4.2.7 (ii) we obtain

$$T_{1,n} \equiv n^{1/2}\frac{(\bar{Y}-\mu)}{\left[\dfrac{\sum_{i=1}^{n}(Y_i-\bar{Y})^2}{n-1}\right]^{1/2}} \xrightarrow{d} N(0,1), \tag{4.9}$$

Hence we see that $T_{1,n}$ is the classical one–sample T–statistic. Clearly the above statistics $T^*_{1,n}$ and $T_{1,n}$ can be used to make inference about the common mean μ (at least asymptotically) even under a heteroskedastic model. Also we notice that

$$\begin{aligned} E(Y_i-\bar{Y})^2 &= E\{(1-\tfrac{1}{n})Y_i - \sum_{j\neq i} Y_j\}^2 \\ &= \mathrm{var}\{(1-\tfrac{1}{n})Y_i - \tfrac{1}{n}\sum_{j\neq i} Y_j\} \\ &= \left[\frac{(n-1)^2\sigma_i^2 + \sum_{j\neq i}\sigma_j^2}{n^2}\right], \end{aligned}$$

and therefore

$$\begin{aligned} E\left[\frac{\sum_{i=1}^{n}(Y_i-\bar{Y})^2}{n-1}\right] &= \frac{\sum_{i=1}^{n}\sigma_i^2}{n} \\ &= \mathrm{var}[n^{1/2}(\bar{Y}-\mu)]. \end{aligned}$$

The above is true for any n. This is the finite sample "*safeness*" property we discussed in Section 3.4 of Chapter 3, that we already knew to be true for the one–sample T–statistic; see for example Cressie (1982). We do expect this since $\hat{V}_{n,mw}$ is also the MINQU estimator of $\bar{V}_n$, proposed by Rao (1970). This can be seen as follows. In the context of this problem we notice that

$$R = \begin{bmatrix} (1-\frac{1}{n}) & \frac{1}{n} & \cdots & \frac{1}{n} \\ \frac{1}{n} & (1-\frac{1}{n}) & \cdots & \frac{1}{n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{n} & \frac{1}{n} & \cdots & (1-\frac{1}{n}) \end{bmatrix},$$

and

$$R^{-1} = \begin{bmatrix} a & b & \cdots & b \\ b & a & \cdots & b \\ \vdots & \vdots & \ddots & \vdots \\ b & b & \cdots & a \end{bmatrix},$$

where $a = \frac{(n^2-n-1)}{(n-1)(n-2)}$ and $b = \frac{-1}{(n-1)(n-2)}$.

Therefore

$$R^{-1}\underline{\epsilon}^2 = \begin{bmatrix} \frac{n}{(n-2)}(Y_1-\bar{Y})^2 - \frac{S^2}{(n-2)} \\ \frac{n}{(n-2)}(Y_2-\bar{Y})^2 - \frac{S^2}{(n-2)} \\ \vdots \\ \frac{n}{(n-2)}(Y_n-\bar{Y})^2 - \frac{S^2}{(n-2)} \end{bmatrix},$$

where

$$S^2 = \frac{\sum_{i=1}^{n}(Y_i-\bar{Y})^2}{(n-1)}.$$

Thus the MINQU estimator of $\bar{V}_n = \frac{\sum_{i=1}^{n}\sigma_i^2}{n}$ is given by $\frac{1}{n}\underline{1}'R^{-1}\underline{\hat{\epsilon}}^2$ where $\underline{1}$ is a vector of ones. After a small simplification one can show that this estimator is exactly equal to $\hat{V}_{n,mw}$.

Now suppose we ask "Can we approximate the finite sample distribution of $T_{1,n}$ by a t—distribution with some equivalent degrees of freedom (e.d.f.)?" This question was answered in a more general context by Cressie (1982) who gave a formula for the e.d.f valid to the O(1). We discussed his results in Section 3.4 of Chapter 3. We will go through the derivation of e.d.f. in this particular situation. Obviously, this is an impossible question to answer unless we make an additional assumption on the distribution of ϵ_i's. Henceforth, in this section we shall assume that the ϵ_i's follow the normal probability law (i.e., $\epsilon_i \sim N(0,\sigma_i^2)$. Now we can write

$$T_{1,n} = n^{1/2} \frac{(\bar{Y}-\mu)}{\left[\dfrac{\sum_{i=1}^{n}(Y_i-\bar{Y})^2}{n-1}\right]^{1/2}}$$

$$= \frac{n^{1/2}(\bar{Y}-\mu)/\left(\dfrac{\sum_{i=1}^{n}\sigma_i^2}{n}\right)^{1/2}}{\left\{\left(\sum_{i=1}^{n}\epsilon_i'^2\right)/\left(\sum_{i=1}^{n}\sigma_i^2\right)\right\}^{1/2}}, \tag{4.10}$$

where $\hat{\epsilon}_i' = \left(\frac{n}{n-1}\right)^{1/2}(Y_i-\bar{Y}) \quad (i=1,2,\ldots,n)$.

We see that the numerator in (4.10) has a standard normal distribution, and after a small calculation it is seen to be uncorrelated with the denominator. Now for $T_{1,n}$ to follow an approximate t—distribution with f_1 degrees of freedom we should have

$$\frac{\left(\sum_{i=1}^{n}\epsilon_i'^2\right)}{\left(\sum_{i=1}^{n}\sigma_i^2\right)} \sim \frac{\chi_{f_1}^2}{f_1},$$

at least approximately.

Therefore to find f_1 we could match the second moments (first moments are already matched). This approach of approximating distributions was first suggested by Smith (1936) and Satterthwaite (1946) who used it to approximate distribution of estimates of variance components. After some algebra, one can show that

$$E\left[\frac{\sum_{i=1}^{n}\epsilon_i'^2}{\sum_{i=1}^{n}\sigma_i^2}\right]^2 = \frac{n^2}{(n-1)^2}\left[\frac{2(n-2)}{n}\ \frac{\sum_{i=1}^{n}\sigma_i^4}{\left(\sum_{i=1}^{n}\sigma_i^2\right)^2} + \frac{(n^2-2n+3)}{n^2}\right].$$

Now setting the above right hand side equal to $E\left[\frac{\chi^2_{f_1}}{f_1}\right]^2 = 1+\frac{2}{f_1}$ and solving for f_1 we obtain

$$f_1 = 2\,\frac{(n-1)^2}{n^2}\left[\frac{2(n-2)}{n}\ \frac{\sum_{i=1}^{n}\sigma_i^4}{\left(\sum_{i=1}^{n}\sigma_i^2\right)^2} + \frac{2}{n^2}\right]^{-1}. \tag{4.11}$$

We immediately notice that if all σ_i^2 's are equal then $T_{1,n}$ follows exactly a t−distribution with $(n-1)$ degrees of freedom under the normality assumption of the ϵ_i's, and we also obtain from (4.11), $f_1 = (n-1)$ as expected. Now to obtain an upper bound for f_1, let ς be a random variable which assumes the values σ_i^2 with probability $\frac{1}{n}$; $i = 1,2,\ldots,n$. Then we can easily see that $E(\varsigma) = \left[\sum_{i=1}^{n}\sigma_i^2\right]/n$ and $E(\varsigma^2) = \left[\sum_{i=1}^{n}\sigma_i^4\right]/n$. Therefore by Jensen's inequality

$$E(\varsigma^2) \geq [E(\varsigma)]^2,$$

and hence

$$\frac{\sum_{i=1}^{n}\sigma_i^4}{\left(\sum_{i=1}^{n}\sigma_i^2\right)^2} \geq \frac{1}{n}.$$

Therefore, from (4.11), it is apparent that

$$f_1 \leq (n-1). \tag{4.12}$$

Now, in order to obtain a lower bound for the equivalent degrees of freedom f_1, we shall the Theorem 3.3.1.

Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed with mean μ^* and variance 1. Consider the weighted unbiased estimator $\bar{X}_w$ of μ^* where

$$\bar{X}_w = \sum_{i=1}^{n} w_i X_i,$$

and

$$w_i = \frac{\sigma_i^2}{\sum_{j=1}^{n} \sigma_j^2}; \qquad (j = 1,2,\ldots,n).$$

Then

$$\operatorname{var}(\bar{X}_w) = \frac{\sum_{i=1}^{n} \sigma_i^4}{\left(\sum_{i=1}^{n} \sigma_i^2\right)^2}.$$

Also $\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$ is the optimally weighted unbiased estimator of μ^* and $\operatorname{var}(\bar{X}) = \frac{1}{n}$. Therefore from the Kantorovich inequality (see Theorem 3.3.1 in Chapter 3) we have

$$\frac{\operatorname{var}(\bar{X}_w)}{\operatorname{var}(\bar{X})} \leq \frac{(Q+1)^2}{4Q},$$

where

$$Q = \frac{\max\{\sigma_i^2;\ i = 1,2,\ldots,n\}}{\min\{\sigma_i^2;\ i = 1,2,\ldots,n\}}.$$

i.e.,

$$\frac{\sum_{i=1}^{n} \sigma_i^4}{\left(\sum_{i=1}^{n} \sigma_i^2\right)^2} \leq \frac{1}{n}\,\frac{(Q+1)^2}{4Q}.$$

Therefore from (4.11)

$$f_1 \geq (n-1)^2 \left[(n-2)\frac{(Q+1)^2}{4Q} + 1 \right]^{-1}. \tag{4.13}$$

Now, combining (4.12) and (4.13) we obtain

$$(n-1)^2 \left[(n-2)\frac{(Q+1)^2}{4Q} + 1 \right]^{-1} \leq f_1 \leq (n-1). \tag{4.14}$$

In most practical situations, the experimenter has some idea about the unknown quantity Q from previous studies or a pilot experiment. Thus one can obtain sensible bounds for the e.d.f. f_1. The degrees of freedom f_1 we obtained above reflects the precision of the estimated variance of $\mathrm{var}(\hat{\mu}) = \mathrm{var}(\bar{Y})$. In the homoskedastic error variance situation, under the normality assumption of the errors we obtain $f_1 = (n-1)$. When the homoskedasticity assumption is relaxed, since we allow unequal error variances we lose some degrees of freedom due to the fact that the σ_i^2's are unknown. The worst one can do is to use the minimum e.d.f. given by (4.14). This minimum e.d.f. certainly depends on the structure of the heteroskedasticity through the quantity Q. The plots of Figure 4.1 show the effect of Q on the minimum e.d.f. for sample sizes $n = 5, 10, 20$, and 50. These plots suggests that for mild heterogeneity among the error variance (e.g., $Q < 1.5$) and for small sample sizes (e.g., $n \leq 20$), we do not lose much in degrees of freedom, i.e., minimum e.d.f. is almost equal to $(n-1)$. As Q increases the drop in the minimum e.d.f. becomes rapid as the sample size increase. It can be shown that for large samples, when Q increases to around 5.8 we lose almost half of the degrees of freedom if we use the minimum e.d.f. When a value of Q is not available to the experimenter one can estimate the e.d.f. by using the residuals $\hat{\varepsilon}_i'$'s instead of σ_i's in the formula given by (4.11).

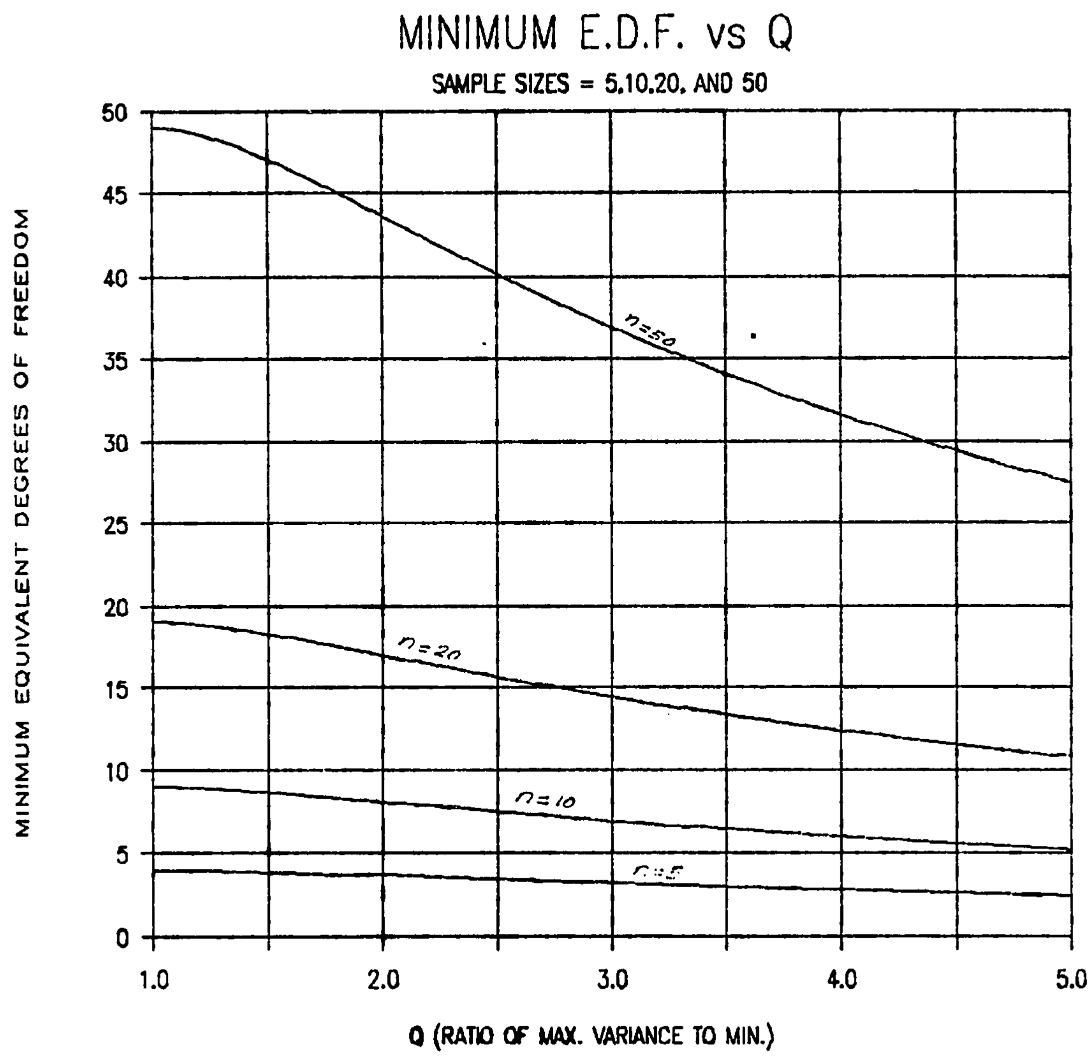


Figure 4.1. Minimum equivalent degrees of freedom vs Q

4.4. Two—Sample Problem

We shall first state the problem briefly. Let the independent observations $\{Y_{ij}\}$ be such that

$$\frac{Y_{ij}-\mu_i}{\sigma_{ij}} \sim G \qquad (i=1,2\ ;\ j=1,2,\ldots,n_i), \tag{4.15}$$

where G is the standard normal cumulative distribution function (c.d.f.) or some other c.d.f. with mean 0 and variance 1. This model was proposed by Cressie and Whitford (1986).

It is a common problem in statistics to perform a test for the equality of means against two sided or one sided alternatives. When the σ_{ij}'s are equal and $G \equiv \Phi$ one could use the classical t—test. When $\sigma_{ij}=\sigma_i$ $(i=1,2)$, $\sigma_1 \neq \sigma_2$ and $G \equiv \Phi$ the problem of comparing means is called the Behrens—Fisher problem. This problem does not have a universally accepted solution. A survey of various proposed solutions to the Behrens—Fisher problem and their power characteristics is given by Scheffe (1970). The most commonly used solution to this problem is given by Welch (1937).

When the independent observations satisfy (4.15) above, it is easy to identify this as a special case of White's model discussed in Section 4.2; i.e., write

$$Y_{ij} = \mu_i + \epsilon_{ij}, \tag{4.16}$$

where $E(\epsilon_{ij})=0$ and $E(\epsilon_{ij}^2)=\sigma_{ij}^2$ $(i=1,2;\ j=1,2,\ldots,n_i)$.

This can be written in matrix notation as

$$\underline{Y} = X\underline{\beta}_o + \underline{\epsilon},$$

where $E(\underline{\epsilon}) = \underline{0}$, $\mathrm{var}(\underline{\epsilon}) = \mathrm{diag}(\sigma_{11}^2, \sigma_{12}^2, \ldots, \sigma_{2n_2}^2)$, $\sigma_{ij}^2 < \infty$; for all i, j,

$$\underline{Y} = \begin{bmatrix} Y_{11} \\ Y_{12} \\ \vdots \\ Y_{2n_2} \end{bmatrix}_{(n_1+n_2)\times 1},$$

$$X = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ \vdots & \vdots \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ \vdots & \vdots \\ 0 & 1 \end{bmatrix}_{(n_1+n_2)\times 2},$$

$$\underline{\beta}_o = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix},$$

and

$$\underline{\epsilon} = \begin{bmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \vdots \\ \epsilon_{2n_2} \end{bmatrix}_{(n_1+n_2)\times 2},$$

Now in this situation, we can see that the ordinary least squares estimator of $\underline{\beta}_o$ is given by

$$\hat{\underline{\beta}}_n = \begin{bmatrix} \hat{\mu}_1 \\ \hat{\mu}_2 \end{bmatrix} = \begin{bmatrix} \bar{Y}_1 \\ \bar{Y}_2 \end{bmatrix},$$

and in the notation we introduced in Section 4.2 that

$$\bar{M}_n = \frac{X'X}{n}$$

$$= \begin{bmatrix} n_1/n & 0 \\ 0 & n_2/n \end{bmatrix},$$

where $n = n_1 + n_2$,

$$\hat{V}_{n,w} = \frac{1}{n} \begin{bmatrix} \sum_{j=1}^{n_1} (Y_{1j} - \bar{Y}_1)^2 & 0 \\ 0 & \sum_{j=1}^{n_2} (Y_{2j} - \bar{Y}_2)^2 \end{bmatrix},$$

and

$$\hat{V}_{n,mw} = \frac{1}{n} \begin{bmatrix} \frac{n_1}{(n_1 - 1)} \sum_{j=1}^{n_1} (Y_{1j} - \bar{Y}_1)^2 & 0 \\ 0 & \frac{n_2}{(n_2 - 1)} \sum_{j=1}^{n_2} (Y_{2j} - \bar{Y}_2)^2 \end{bmatrix},$$

where

$$\bar{Y}_i = \frac{\sum_{j=1}^{n_i} Y_{ij}}{n_i} \quad (i=1,2).$$

Now suppose we wish to test the hypotheses

$$H_O: \mu_1 = \mu_2 \qquad \text{vs} \qquad H_1: \mu_1 \neq \mu_2 .$$

In matrix notation

$$H_O: (\ 1 \quad -1\ ) \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} = 0 \qquad \text{vs} \qquad H_1: (\ 1 \quad -1\ ) \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \neq 0 .$$

We notice here in connection to Theorem 4.2.7 that

$$R = (\ 1 \quad -1\ ), \text{ and } \underline{r} = \underline{0}.$$

Therefore by Theorem 4.2.7 (iii) we obtain

$$\frac{(\bar{Y}_1 - \bar{Y}_2)^2}{\left[\frac{\sum^{n_1}(Y_{1j} - \bar{Y}_1)^2}{n_1(n_1 - 1)} + \frac{\sum^{n_2}(Y_{2j} - \bar{Y}_2)^2}{n_2(n_2 - 1)}\right]} \xrightarrow{d} \chi_1^2 \tag{4.17}$$

Hence we obtain a computable statistic to test H_0 vs H_1, at least asymptotically.

As was discussed in the one—sample situation, we may be interested in approximating the finite sample distribution of the above statistic. To address this, henceforth let us assume that the random error terms are normally distributed. Now we could view (4.16) as

$$\frac{(\bar{Y}_1 - \bar{Y}_2)^2}{\left[\frac{\sum_{j=1}^{n_1}(Y_{1j} - \bar{Y}_1)^2}{n_1(n_1 - 1)} + \frac{\sum_{j=1}^{n_2}(Y_{2j} - \bar{Y}_2)^2}{n_2(n_2 - 1)}\right]} \xrightarrow{d} t_\infty^2 .$$

Thus we could try to approximate the finite sample distribution of

$$T_{2,n} \equiv \frac{(\bar{Y}_1 - \bar{Y}_2)}{\left[\frac{\sum_{j=1}^{n_1}(Y_{1j} - \bar{Y}_1)^2}{n_1(n_1 - 1)} + \frac{\sum_{j=1}^{n_2}(Y_{2j} - \bar{Y}_2)^2}{n_2(n_2 - 1)}\right]^{1/2}} ,$$

by a t—distribution with some equivalent degrees of freedom. Therefore let us suppose that

$$T_{2,n} \sim t_{f_2},$$

at least approximately.

We notice here that the statistic given by $T_{2,n}$ is *safe* even in the finite sample situations. For brevity let us write

$$S_i^2 = \frac{\sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2}{(n_i - 1)} \qquad (i = 1,2).$$

Now we can write $T_{2,n}$ as

$$T_{2,n} = \frac{(\bar{Y}_1 - \bar{Y}_2)}{\left[\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right]^{1/2}}. \tag{4.18}$$

At this point it is interesting to compare T_{2n} to the usual two—sample test statistics T_{2n}^* which is given by

$$T_{2,n}^* = \frac{(\bar{Y}_1 - \bar{Y}_2)}{\left[\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{(n_1 + n_2)(n_1 + n_2 - 2)}\right]^{1/2}}.$$

It can be shown easily that even if $\sigma_{ij}^2 = \sigma_i^2$ for $i = 1,2$ and if $\sigma_1^2 \neq \sigma_2^2$, then $T_{2,n}^*$ converges to a normal distribution, but with variance typically different from 1. Therefore even in this special situation, without some knowledge that $\sigma_1^2 = \sigma_2^2$ one should not try to approximate its distribution by a t—distribution. On the other hand we immediately see that $T_{2,n}$ is safe and by Slutsky's theorem it converges to a standard normal distribution.

After some algebra one can show that the numerator of $T_{2,n}$ given by (4.18) is uncorrelated with the denominator. Hence under normal errors it is quite reasonable to try to approximate the finite sample distribution of $T_{2,n}$ by a t—distribution with some appropriate degrees of freedom. To understand the method of approximating the distribution of $T_{2,n}$ let us rewrite it as follows.

$$T_{2,n} = \frac{(\bar{Y}_1 - \bar{Y}_2) \; / \; D^{1/2}}{\left[\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)/D\right]^{1/2}},$$

where

$$D = \left[\frac{\sum_{j=1}^{n_1} \sigma_{1j}^{\;2}}{n_1^2} + \frac{\sum_{j=1}^{n_2} \sigma_{2j}^2}{n_2^2}\right]. \tag{4.19}$$

Under the assumption that the error terms are normally distributed, we immediately see that the above numerator has a standard normal distribution and it is uncorrelated with the denominator. Thus for $T_{2,n}$ to follow approximately a t—distribution with f_2 degrees of freedom our intuition tells us that the above denominator should behave like a $\left(\chi_{f_2}^2/f_2\right)^{1/2}$; i.e., we should have

$$\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right) \Big/ \left[\frac{\sum_{j=1}^{n_1} \sigma_{1j}^{\;2}}{n_1^2} + \frac{\sum_{j=1}^{n_2} \sigma_{2j}^2}{n_2^2}\right] \sim \frac{\chi_{f_2}^2}{f_2}.$$

Now we can obtain an approximate value for f_2 by matching the second moments (we match second moments since the first moments are already matched), i.e., to obtain f_2 we should solve

$$\left[\frac{\sum_{j=1}^{n_1} \sigma_{1j}^{\;2}}{n_1^2} + \frac{\sum_{j=1}^{n_2} \sigma_{2j}^2}{n_2^2}\right]^{-2} E\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2 = 1 + \frac{2}{f_2}. \tag{4.20}$$

Now

$$E\left(\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}\right)^2 = E\left(\frac{S_1^4}{n_1^2}+2\,\frac{S_1^2}{n_1}\frac{S_2^2}{n_2}+\frac{S_2^4}{n_2^2}\right), \tag{4.21}$$

and

$$E(S_i^4) = \frac{1}{(n_i-1)^2}E\left(\sum_{j=1}^{n_i}(Y_{ij}-\bar{Y}_i)^2\right)^2 \qquad (i=1,2).$$

Therefore by similar reasoning to the one—sample problem we obtain

$$E\left(\sum_{j=1}^{n_i}(Y_{ij}-\bar{Y}_i)^2\right)^2 = \frac{2(n_i-2)}{n_i}\sum_{j=1}^{n_i}\sigma_{ij}^4 + \frac{(n_i^2-2n_i+3)}{n_i^2}\left(\sum_{j=1}^{n_i}\sigma_{ij}^2\right)^2$$

$$(i=1,2). \tag{4.22}$$

Also notice that

$$E(S_i^2) = \frac{\sum_{j=1}^{n_i}\sigma_{ij}^2}{n_i} \qquad (i=1,2). \tag{4.23}$$

Substituting (4.22) and (4.23) in (4.21) we obtain

$$E\left(\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}\right)^2 = \frac{1}{n_1^2(n_1-1)^2}\left[\frac{2(n_1-2)}{n_1}\sum_{j=1}^{n_1}\sigma_{1j}^4 + \frac{(n_1^2-2n_1+3)}{n_1^2}\left(\sum_{j=1}^{n_1}\sigma_{1j}^2\right)^2\right]$$

$$+\frac{1}{n_2^2(n_2-1)^2}\left[\frac{2(n_2-2)}{n_2}\sum_{j=1}^{n_2}\sigma_{2j}^4 + \frac{(n_2^2-2n_2+3)}{n_2^2}\left(\sum_{j=1}^{n_2}\sigma_{2j}^2\right)^2\right]$$

$$+\frac{2\left(\sum_{j=1}^{n_1}\sigma_{1j}^2\right)\left(\sum_{j=1}^{n_2}\sigma_{2j}^2\right)}{n_1^2\,n_2^2}.$$

Simplifying further we obtain

$$E\left(\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}\right)^2 = \frac{2(n_1-2)}{n_1^3(n_1-1)^2}\sum_{j=1}^{n_1}\sigma_{1j}^4 + \frac{2\left(\sum_{j=1}^{n_1}\sigma_{1j}^2\right)^2}{n_1^4(n_1-1)^2}$$

$$+\frac{2(n_2-2)}{n_2^3(n_2-1)^2}\sum_{j=1}^{n_2}\sigma_{2j}^4 + \frac{2\left(\sum_{j=1}^{n_2}\sigma_{2j}^2\right)^2}{n_2^4(n_2-1)^2}$$

$$+\left[\frac{\sum_{j=1}^{n_1}\sigma_{1j}^2}{n_1^2}+\frac{\sum_{j=1}^{n_2}\sigma_{2j}^2}{n_2^2}\right]^2.$$

Therefore

$$\frac{E\left(\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}\right)^2}{D^2} = 1+\frac{2}{D^2}\left\{\frac{(n_1-2)}{n_1^3(n_1-1)^2}\sum_{j=1}^{n_1}\sigma_{1j}^4+\frac{\left(\sum_{j=1}^{n_1}\sigma_{1j}^2\right)^2}{n_1^4(n_1-1)^2}\right.$$

$$\left.+\frac{(n_2-2)}{n_2^3(n_2-1)^2}\sum_{j=1}^{n_2}\sigma_{2j}^4+\frac{\left(\sum_{j=1}^{n_2}\sigma_{2j}^2\right)^2}{n_2^4(n_2-1)^2}\right\},$$

where D is defined by (4.19).

Now from (4.20), we obtain the equivalent degrees of freedom for $T_{2,n}$; namely

$$f_2=\left(\frac{\sum_{j=1}^{n_1}\sigma_{1j}^2}{n_1^2}+\frac{\sum_{j=1}^{n_2}\sigma_{2j}^2}{n_2^2}\right)^2\left(\frac{n_1(n_1-2)\sum_{j=1}^{n_1}\sigma_{1j}^4+\left(\sum_{j=1}^{n_1}\sigma_{1j}^2\right)^2}{n_1^4(n_1-1)^2}+\right.$$

$$\left.\frac{n_2(n_2-2)\sum_{j=1}^{n_2}\sigma_{2j}^4+\left(\sum_{j=1}^{n_2}\sigma_{2j}^2\right)^2}{n_2^4(n_2-1)^2}\right)^{-1}. \qquad (4.24)$$

As we saw in the one—sample problem in the previous section one can show using Jensen's inequality that

$$\frac{\sum_{j=1}^{n_i} \sigma_{ij}^4}{\left(\sum_{j=1}^{n_i} \sigma_{ij}^2\right)^2} \geq \frac{1}{n_i} \qquad (i=1,2).$$

Hence it is apparent from (4.24) that

$$f_2 \leq \left(\frac{\sum_{j=1}^{n_1} \sigma_{1j}^2}{n_1^2} + \frac{\sum_{j=1}^{n_2} \sigma_{2j}^2}{n_2^2}\right)^2 \left[\frac{\left(\sum_{j=1}^{n_1} \sigma_{1j}^4\right)^2}{n_1^4(n_1-1)} + \frac{\left(\sum_{j=1}^{n_2} \sigma_{2j}^4\right)^2}{n_2^4(n_2-1)}\right]^{-1},$$

and using calculus one can easily show that

$$\left(\frac{\sum_{j=1}^{n_1} \sigma_{1j}^2}{n_1^2} + \frac{\sum_{j=1}^{n_2} \sigma_{2j}^2}{n_2^2}\right)^2 \left[\frac{\left(\sum_{j=1}^{n_1} \sigma_{1j}^4\right)^2}{n_1^4(n_1-1)} + \frac{\left(\sum_{j=1}^{n_2} \sigma_{2j}^4\right)^2}{n_2^4(n_2-1)}\right]^{-1} \leq (n_1+n_2-2).$$

Hence

$$f_2 \leq (n_1+n_2-2). \tag{4.25}$$

Now we will try to obtain a sensible lower bound for f_2. As we discussed in the one—sample problem we can show that

$$\frac{\sum_{j=1}^{n_1} \sigma_{1j}^4}{\left(\sum_{j=1}^{n_1} \sigma_{1j}^2\right)^2} \leq \frac{1}{n_1} \frac{(Q_1+1)^2}{4Q_1}, \tag{4.26}$$

and

$$\frac{\sum_{j=1}^{n_2} \sigma_{2j}^4}{\left(\sum_{j=1}^{n_2} \sigma_{2j}^2\right)^2} \leq \frac{1}{n_2} \frac{(Q_2+1)^2}{4Q_2}, \tag{4.27}$$

where

$$Q_i = \frac{\max\{\sigma_{ij}^2;\ j=1,2,\ldots,n_i\}}{\min\{\sigma_{ij}^2;\ j=1,2,\ldots,n_i\}}, \quad (i=1,2). \tag{4.28}$$

Substituting (4.26) and (4.27) in (4.24) yields

$$f_2 \geq \frac{(A+B)^2}{(A^2C+B^2D)}, \tag{4.29}$$

where

$$A = \frac{\sum_{j=1}^{n_1} \sigma_{1j}^2}{n_1^2}, \tag{4.30}$$

$$B = \frac{\sum_{j=1}^{n_2} \sigma_{2j}^2}{n_2^2}, \tag{4.31}$$

$$C = \frac{1}{(n_1-1)^2}\left[(n_1-2)\frac{(Q_1+1)^2}{4Q_1} + 1\right], \tag{4.32}$$

and

$$D = \frac{1}{(n_2-1)^2}\left[(n_2-2)\frac{(Q_2+1)^2}{4Q_2} + 1\right], \tag{4.33}$$

where Q_1 and Q_2 are given by (4.29).

One can rewrite the right hand side of (4.30) as follows:

$$\frac{(A+B)^2}{(A^2C+B^2D)} = \frac{1}{\left(\frac{A}{A+B}\right)^2 C + \left(\frac{B}{A+B}\right)^2 D}.$$

Let X_1 and X_2 be independent random variables with common mean μ^* and variances C and D respectively. Then the optimally weighted unbiased estimator of μ^* is given by

$$\hat{\mu}_O^* = \frac{\frac{1}{C}X_1 + \frac{1}{D}X_2}{\frac{1}{C} + \frac{1}{D}},$$

and

$$\mathrm{var}(\hat{\mu}_O^*) = \frac{1}{(\frac{1}{C} + \frac{1}{D})}. \tag{4.34}$$

Now consider the weighted unbiased estimator of μ^*, given by

$$\hat{\mu}_W^* = \frac{A}{(A+B)}X_1 + \frac{B}{(A+B)}X_2.$$

Then

$$\mathrm{var}(\hat{\mu}_W^*) = \frac{A^2}{(A+B)^2}C + \frac{B^2}{(A+B)^2}D. \tag{4.35}$$

Therefore by the Kantorovich theorem (see Theorem 3.3.1) we obtain

$$\frac{\mathrm{var}(\hat{\mu}_W^*)}{\mathrm{var}(\hat{\mu}_O^*)} \le \frac{(Q+1)^2}{4Q}, \tag{4.36}$$

where

$$Q = \frac{\max\{\frac{A}{(A+B)}C\ ,\ \frac{B}{(A+B)}D\}}{\min\{\frac{A}{(A+B)}C\ ,\ \frac{B}{(A+B)}D\}},$$

i.e.,

$$Q = \frac{\max\{AC,BD\}}{\min\{AC,BD\}}, \tag{4.37}$$

where A, B, C, and D are given by (4.30) — (4.33) respectively.

Therefore substituting (4.34) and (4.35) in (4.36) we obtain

$$\frac{A^2}{(A+B)^2}C + \frac{B^2}{(A+B)^2}D \leq \left[\frac{1}{C}+\frac{1}{D}\right]\frac{(Q+1)^2}{4Q},$$

i.e.,

$$\frac{(A+B)^2}{(A^2C+B^2D)} \geq \frac{4QCD}{(C+D)(Q+1)^2},$$

and by (4.29)

$$f_2 \geq \frac{4QCD}{(C+D)(Q+1)^2}. \tag{4.38}$$

Combining (4.25) and (4.39) we obtain

$$\frac{4QCD}{(C+D)(Q+1)^2} \leq f_2 \leq (n_1+n_2-2), \tag{4.39}$$

where Q,C, and D are given by (4.37),(4.32), and (4.33) respectively.

The worst one could do in using the e.d.f. f_2 above is to use the minimum e.d.f. given by the inequality (4.39). To obtain the minimum e.d.f. one should know the values Q, C, and D. In other words one should know Q, Q_1, and Q_2. Usually these are available to the experimenter from previous studies or from pilot surveys.

The model (4.15) we considered in this section which was introduced by Cressie and Whitford (1986) is in a more general context. In the usual two-sample problem we assume the homogeneity and normality of all the observations. The next generalization is to assume homogeneity of variances within groups and variances possibly unequal between groups. Under normality this problem of comparing means is called the Behrens-Fisher problem. Another generalization of the usual two sample problem can be obtained as follows. We allow unequal variances within groups, but assume that the average of the variances within a group is an unknown constant which does not vary from group to group, i.e., we assume the Cressie—Whitford model given by (4.15) with

$$\frac{\sum_{j=1}^{n_1} \sigma_{1j}^2}{n_1} = \frac{\sum_{j=1}^{n_2} \sigma_{2j}^2}{n_2}. \tag{4.40}$$

That is, we assume that $n_1 A = n_2 B$ where A and B are given by (4.30) and (4.31). It is worthy commenting about the condition given by (4.40) we imposed on the Cressie—Whitford model. In reality this condition may not hold exactly, but we believe it is reasonable to assume it to hold at least approximately. Therefore from (4.29) we obtain a lower bound for the e.d.f. f_2^* for this special case, i.e.,

$$f_2^* \geq \frac{(n_1 + n_2)^2}{(n_1^2 D + n_2^2 C)}. \tag{4.41}$$

Now combining (4.25) and (4.41) we obtain

$$\frac{(n_1 + n_2)^2}{(n_1^2 D + n_2^2 C)} \leq f_2^* \leq (n_1 + n_2 - 2), \tag{4.42}$$

where C and D are defined by (4.30) and (4.31).

We notice here that the minimum e.d.f. f_2^* given by (4.43) is a function of Q_1 and Q_2 as C depends on Q_1 and D depends on Q_2. In a situation where the observations are generated according to the Cressie—Whitford model with the additional conditions given by (4.40) the worst one can do in testing for equality of means is to use the minimum e.d.f. given by (4.42). We should ask "How many degrees of freedom do we lose if we plan to use the minimum e.d.f. for testing purposes?" In other words how much precision at the most do we have to sacrifice in estimating $\mathrm{var}(\bar{Y}_1 - \bar{Y}_2)$? Figures 4.2—4.5 show the contour plots of minimum e.d.f. for the special cases $n_1 = n_2 = 5, 10, 20$, and 50. A quick inspection of these plots shows that if $Q_1, Q_2 < 1.5$, then we retain almost all the degrees of freedom for $n_i \leq 20$.

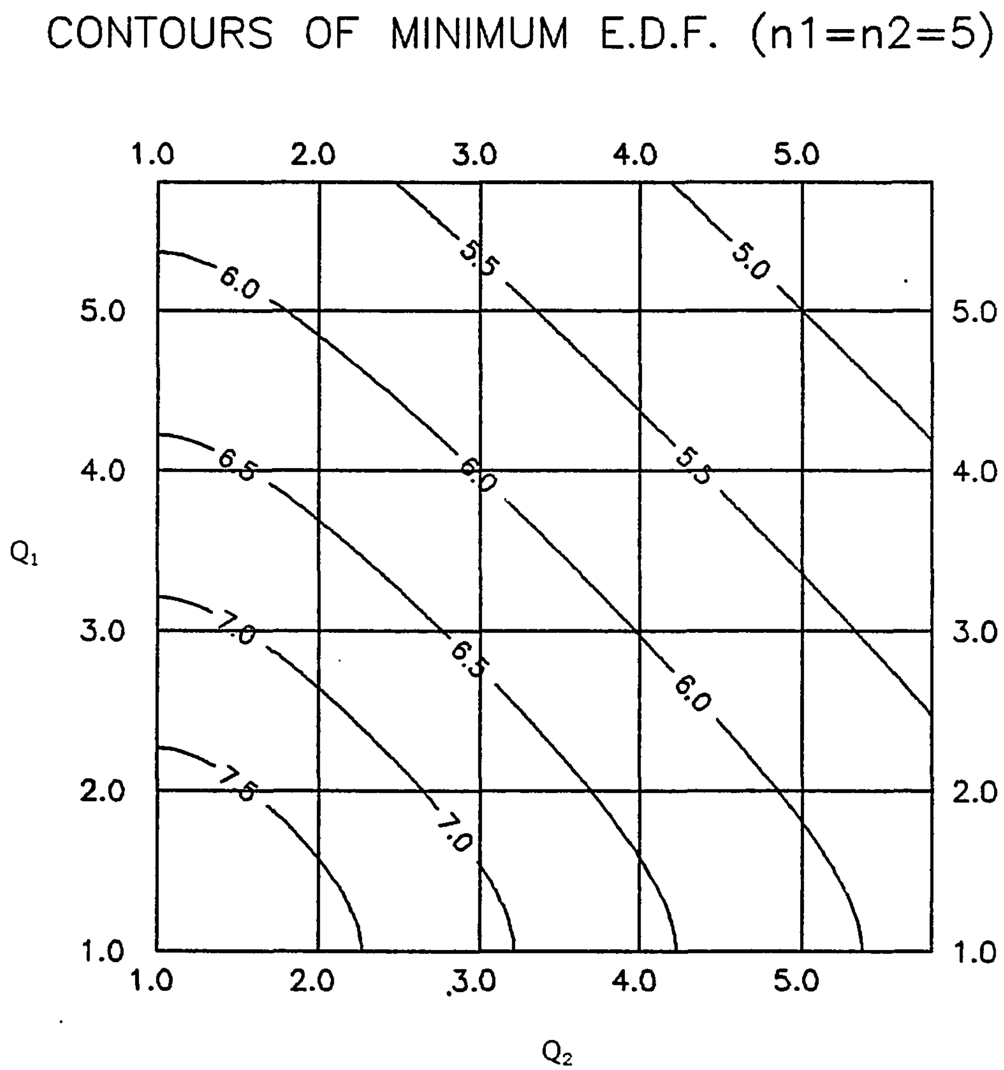


Figure 4.2. Contours of minimum equivalent degrees of freedom, $n_1 = n_2 = 5$

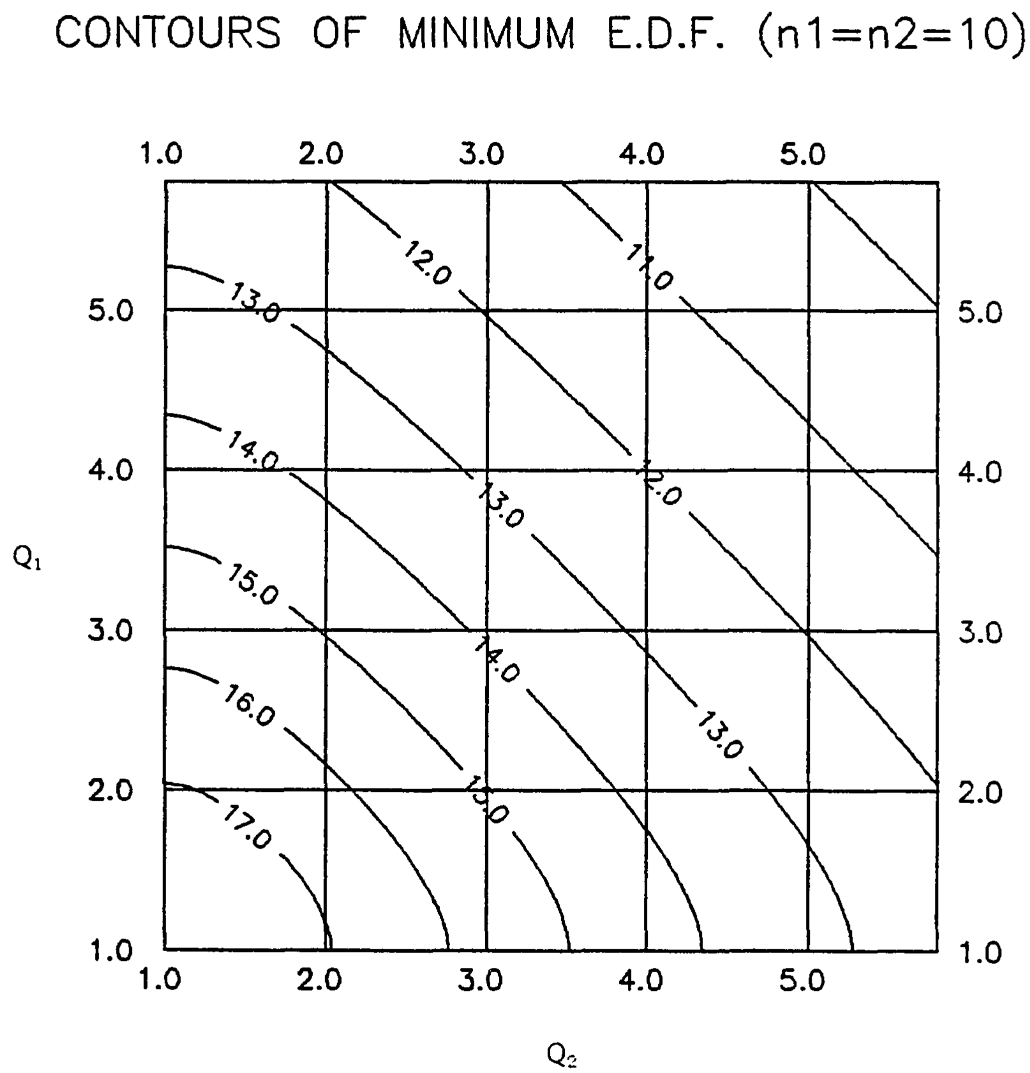


Figure 4.3. Contours of minimum equivalent degrees of freedom, $n_1 = n_2 = 10$

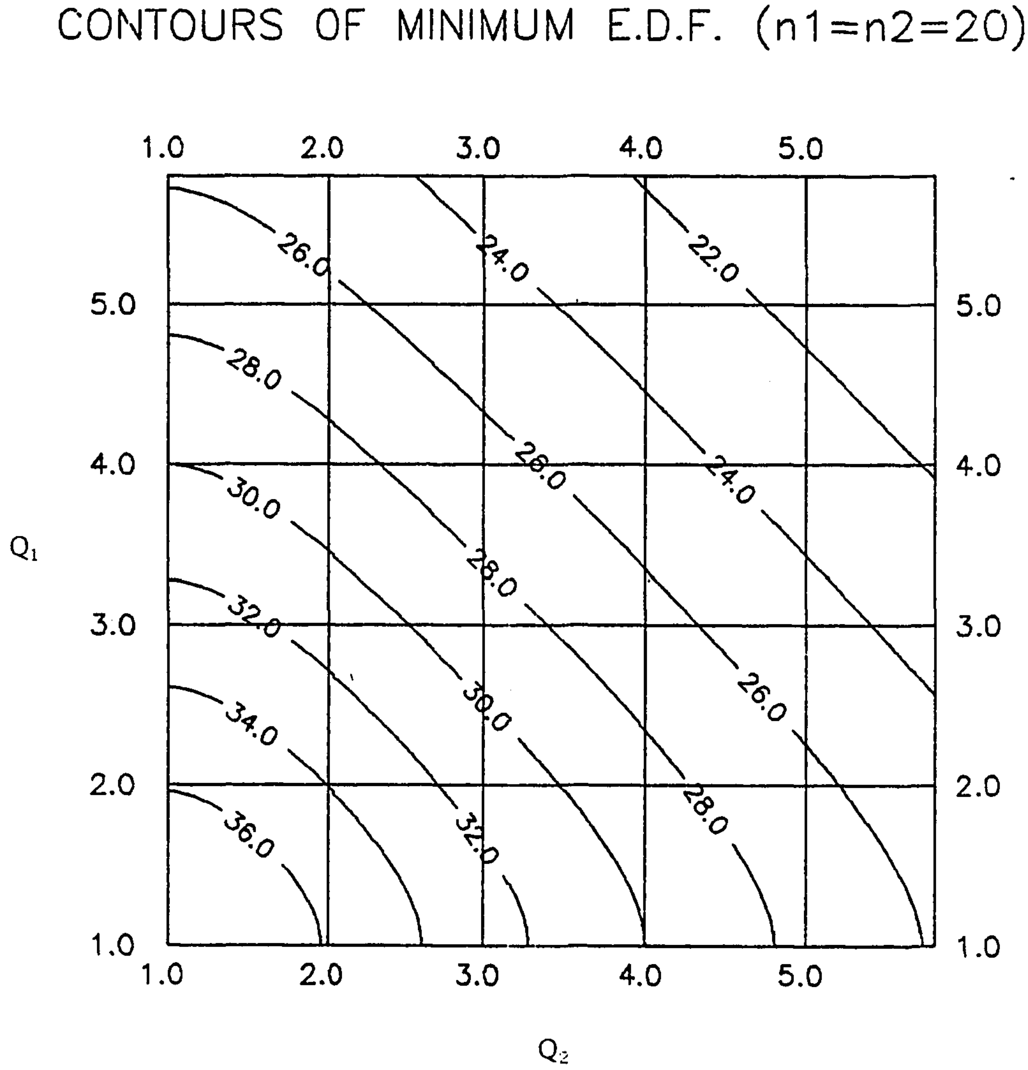


Figure 4.4. Contours of minimum equivalent degrees of freedom, $n_1 = n_2 = 20$

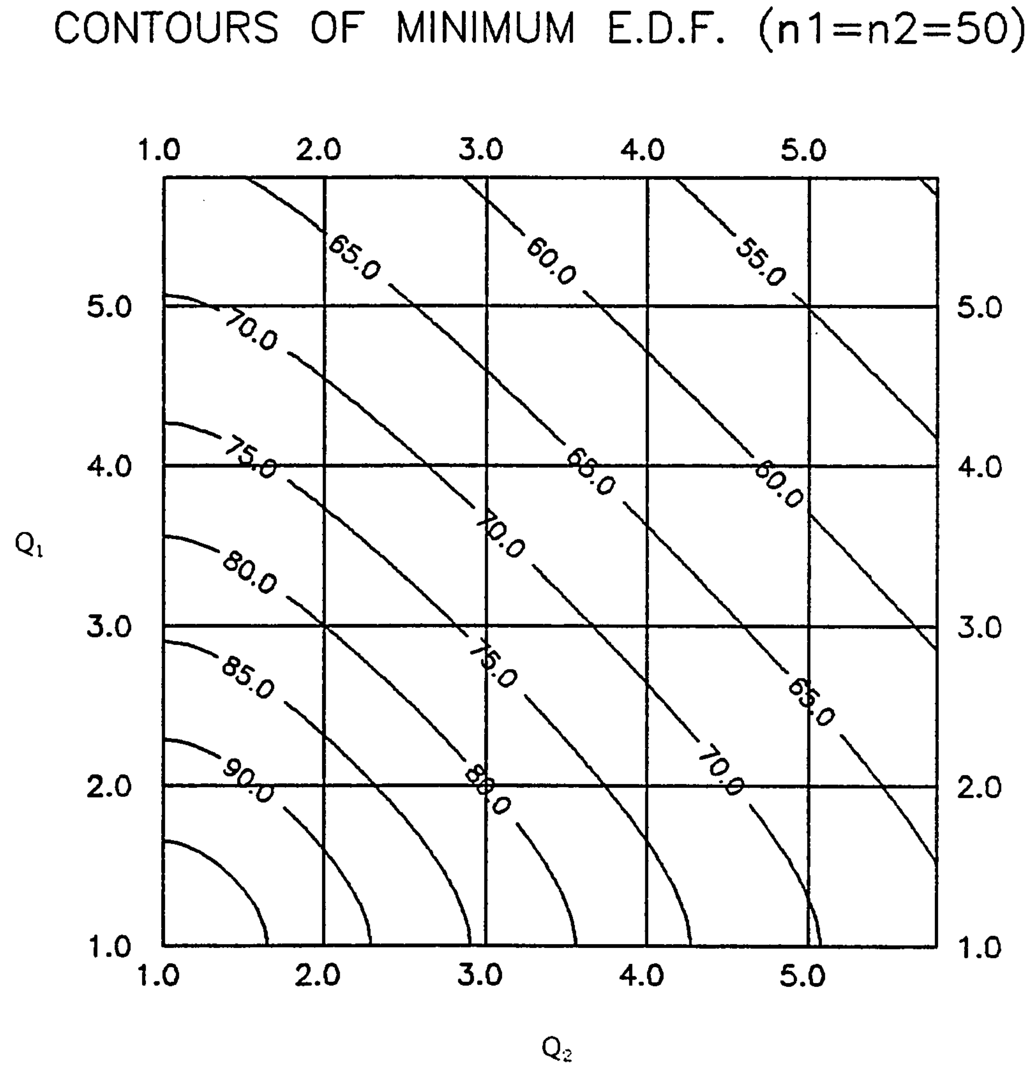


Figure 4.5. Contours of minimum equivalent degrees of freedom, $n_1 = n_2 = 50$

4.5. Simple Linear Regression

In this section we shall be applying White's results that we discussed in Section 4.2 to the simple linear regression problem. First we will look at the simple linear regression problem without intercept and then we will proceed to the case with intercept.

4.5.1. Simple linear regression without intercept

Let

$$Y_i = X_i\beta + \epsilon_i; \quad (i=1,2,\ldots,n), \tag{4.43}$$

where $E(\epsilon_i)=0$, $E(\epsilon_i^2)=\sigma_i^2<\infty$; $(i=1,2,\ldots,n)$ and the ϵ_i's are independent.

In matrix notation one can write

$$\underline{Y} = X\beta + \underline{\epsilon},$$

where

$$\underline{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}_{n \times 1},$$

$$X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix}_{n \times 1},$$

and

$$\underline{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}_{n \times 1} .$$

Then we see that in the notation we introduced in Section 4.2 that

$$X'X = \sum_{i=1}^{n} X_i^2,$$

$$\bar{M}_n = \frac{X'X}{n} = \frac{\sum_{i=1}^{n} X_i^2}{n},$$

and

$$\bar{V}_n = \frac{\sum_{i=1}^{n} X_i^2 \sigma_i^2}{n}.$$

Therefore the ordinary least squares estimator of β is given by

$$\hat{\beta}_n = \frac{\sum_{i=1}^{n} X_i Y_i}{\sum_{i=1}^{n} X_i^2}, \tag{4.44}$$

and

$$\mathrm{var}(\hat{\beta}_n) = \frac{\sum_{i=1}^{n} X_i^2 \sigma_i^2}{\left(\sum_{i=1}^{n} X_i^2 \right)^2}. \tag{4.45}$$

In many situations in statistics where the simple-linear-regression-without-intercept-model is applicable one often wishes to construct confidence intervals for the unknown slope parameter β. We could use Theorem 4.2.7 (ii) to obtain a pivotal quantity and construct the required confidence interval at least asymptotically. Before we apply Theorem 4.2.7 (ii) we need to determine $\hat{V}_{n,mw}$ in the context of this problem.

Now since

$$
\begin{aligned}
H &= X(X'X)^{-1}X' \\
&= \frac{1}{\left(\sum_{i=1}^{n} X_i^2\right)^2}
\begin{bmatrix}
X_1^2 & X_1X_2 & \cdots & X_1X_n \\
X_2X_1 & X_2^2 & \cdots & X_2X_n \\
\vdots & \vdots & \ddots & \vdots \\
X_nX_1 & X_nX_2 & \cdots & X_n^2
\end{bmatrix},
\end{aligned}
$$

we notice immediately that

$$
h_{ii} = \frac{X_i^2}{\left(\sum_{j=1}^{n} X_j^2\right)},
$$

and

$$
1 - h_{ii} = \frac{\left(\sum_{j=1}^{n} X_j^2\right) - X_i^2}{\left(\sum_{j=1}^{n} X_j^2\right)}.
$$

Hence

$$
\begin{aligned}
\hat{V}_{n,mw} &= \frac{1}{n} \sum_{i=1}^{n} (X_i \hat{\epsilon}_i'^2 X_i') \\
&= \frac{1}{n} \sum_{i=1}^{n} X_i \frac{\hat{\epsilon}_i^2}{(1 - h_{ii})} X_i' \\
&= \frac{\left(\sum_{j=1}^{n} X_j^2\right)}{n} \sum_{i=1}^{n} \frac{X_i^2}{\left(\sum_{j=1}^{n} X_j^2 - X_i^2\right)} (Y_i - X_i \hat{\beta}_n)^2 .
\end{aligned}
$$

Recall that $\hat{V}_{n,mw}$ is an estimator of $\bar{V}_n$. It is interesting to notice here that although for one—sample and two—sample problems $\hat{V}_{n,mw}$ estimates $\bar{V}_n$ unbiasedly, when applied to the simple linear regression without intercept problem we shall see that it does not.

First notice that

$$E(Y_i - X_i\hat{\beta}_n) = 0.$$

Therefore

$$E(Y_i - X_i\hat{\beta}_n)^2 = \mathrm{var}(Y_i - X_i\hat{\beta}_n)$$

$$= \mathrm{var}\left(Y_i - X_i \frac{\left(\sum_{j=1}^{n} X_j Y_j\right)}{\left(\sum_{i=1}^{n} X_j^2\right)}\right)$$

$$= \mathrm{var}\left(\left(1 - \frac{X_i^2}{\left(\sum_{i=1}^{n} X_j^2\right)}\right) Y_i - X_i \frac{\left(\sum_{j\neq i} X_j Y_j\right)}{\left(\sum_{i=1}^{n} X_j^2\right)}\right)$$

$$= \left(1 - \frac{X_i^2}{\left(\sum_{i=1}^{n} X_j^2\right)}\right)^2 \sigma_i^2 + X_i^2 \frac{\left(\sum_{j\neq i} X_j^2\sigma_j^2\right)}{\left(\sum_{i=1}^{n} X_j^2\right)^2}$$

$$= \left(1 - \frac{2X_i^2}{\left(\sum_{i=1}^{n} X_j^2\right)}\right)\sigma_i^2 + X_i^2 \frac{\left(\sum_{j=1}^{n} X_j^2\sigma_j^2\right)}{\left(\sum_{i=1}^{n} X_j^2\right)^2} \tag{4.46}$$

Hence

$$E(\hat{V}_{n,mw}) = \frac{\left(\sum_{j=1}^{n} X_j^2\right)}{n} \sum_{i=1}^{n} \frac{X_i^2}{\left(\sum_{j=1}^{n} X_j^2 - X_i^2\right)} E(Y_i - X_i\hat{\beta}_n)^2$$

$$= \frac{\left(\sum_{j=1}^{n} X_j^2\right)}{n} \sum_{i=1}^{n} \frac{X_i^2}{\left(\sum_{j=1}^{n} X_j^2 - X_i^2\right)} \left(\left(1 - \frac{2X_i^2}{\left(\sum_{i=1}^{n} X_j^2\right)}\right)\sigma_i^2 + X_i^2 \frac{\left(\sum_{j=1}^{n} X_j^2\sigma_j^2\right)}{\left(\sum_{i=1}^{n} X_j^2\right)^2}\right).$$

From the right hand side of the equation above we readily see that unless $X_i \equiv 1$ or all σ_i^2's are equal, $E(\hat{V}_{n,mw}) \neq \bar{V}_n$, and hence $\hat{V}_{n,mw}$ is not an unbiased estimator of $\bar{V}_n$ in general. A quick inspection of the right hand side of (4.46) suggests the use of the following estimator (which we call $\hat{V}_{n,unb}$) for $\bar{V}_n$:

$$\hat{V}_{n,unb} = \left[1 + \frac{1}{\left(\sum_{j=1}^{n} X_j^2\right)} \sum_{i=1}^{n} \frac{X_i^4}{\left(\left(\sum_{j=1}^{n} X_j^2\right) - 2X_i^2\right)}\right]^{-1} \cdot \left[\frac{\left(\sum_{j=1}^{n} X_j^2\right)}{n} \sum_{i=1}^{n} \frac{X_i^2}{\left(\left(\sum_{j=1}^{n} X_j^2\right) - 2X_i^2\right)} (Y_i - X_i\hat{\beta}_n)^2\right]. \quad (4.47)$$

It is not difficult to show that $\hat{V}_{n,unb}$ is an unbiased estimator of $\bar{V}_n$ and hence we obtain an unbiased estimator of $\text{var}(\hat{\beta}_n)$. That is, an unbiased estimator of $\text{var}(\hat{\beta}_n)$, which we shall call $\tilde{V}_{n,unb}$, is given by

$$\tilde{V}_{n,unb} = \left[\left(\sum_{j=1}^{n} X_j^2\right) + \sum_{i=1}^{n} \frac{X_i^4}{\left(\left(\sum_{j=1}^{n} X_j^2\right) - 2X_i^2\right)}\right]^{-1} \cdot \left[\sum_{i=1}^{n} \frac{X_i^2}{\left(\left(\sum_{j=1}^{n} X_j^2\right) - 2X_i^2\right)} (Y_i - X_i\hat{\beta}_n)^2\right]. \quad (4.48)$$

As an immediate application of the above estimator, consider the one—sample problem we discussed in Section 4.3. There we considered the model

$$Y_i = \mu + \epsilon_i; \quad (i=1,2,\ldots,n). \tag{4.49}$$

Let $\{w_i \,;\, i = 1,2,\ldots,n\}$ be such that $w_i > 0$ and $\sum_{i=1}^{n} w_i = 1$. Therefore from (4.49) we obtain

$$w_i^{1/2} Y_i = w_i^{1/2} \mu + w_i^{1/2} \epsilon_i; \quad (i=1,2,\ldots,n). \tag{4.50}$$

One recognizes that (4.50) falls in the framework of (4.43). Therefore the ordinary least squares estimator of μ under the model (4.43) is given by

$$\bar{Y}_w = \sum_{i=1}^{n} w_i Y_i,$$

is a weighted linear unbiased estimator of μ. Now applying (4.48) one can obtain an unbiased estimator of $\mathrm{var}(\bar{Y}_w)$, namely

$$\hat{\mathrm{var}}(\bar{Y}_w) = \left[1 + \sum_{i=1}^{n} \frac{w_i^2}{(1 - 2w_i^2)}\right]^{-1} \left[\sum_{i=1}^{n} \frac{w_i^2}{(1 - 2w_i^2)} (Y_i - \bar{Y}_w)^2\right].$$

We recognize the unbiased estimator of $\mathrm{var}(\bar{Y}_w)$ above as exactly $S^2_{w,\delta}$ given by Cressie (1982), which we discussed under safe T—statistics in Chapter 3 (see equation (3.37)).

Returning to the original problem, at this point we have three estimators of $\bar{V}_n$, namely $\hat{V}_{n,w}$ which was suggested by White (1980a), $\hat{V}_{n,mw}$ which was suggested by MacKinnon and White (1985) and $\hat{V}_{n,unb}$ given by (4.48). Hence we obtain three estimators of $\mathrm{var}(\hat{\beta}_n)$:

$$\tilde{V}_{n,w} = \frac{1}{n} \frac{1}{\left(\sum_{i=1}^{n} X_i^2\right)^2} \sum_{i=1}^{n} X_i^2 (Y_i - X_i \hat{\beta}_n)^2, \tag{4.51}$$

$$\tilde{V}_{n,mw} = \frac{1}{n} \frac{1}{\left(\sum_{i=1}^{n} X_i^2\right)} \sum_{i=1}^{n} \frac{X_i^2}{\left(\sum_{j=1}^{n} X_j^2 - X_i^2\right)} (Y_i - X_i \hat{\beta}_n)^2, \tag{4.52}$$

and $\tilde{V}_{n,unb}$ given by (4.48).

Under the assumption of homoskedasticity, in the model (4.15) an estimator of $\mathrm{var}(\hat{\beta}_n)$ is available, which we call the "usual" estimator, $\tilde{V}_{n,usu}$:

$$\tilde{V}_{n,usu} \equiv \frac{1}{(n-1)} \frac{\sum_{i=1}^{n} (Y_i - X_i \hat{\beta}_n)^2}{\left(\sum_{i=1}^{n} X_i^2\right)}. \tag{4.53}$$

As we noted earlier, $\tilde{V}_{n,w}$ and $\tilde{V}_{n,mw}$ are not unbiased estimators of $\mathrm{var}(\hat{\beta}_n)$ in general, but they are asymptotically unbiased. Furthermore, $\tilde{V}_{n,mw}$, $\tilde{V}_{n,unb}$, and $\tilde{V}_{n,usu}$ are unbiased under the homoskedastic model and $\tilde{V}_{n,unb}$ is unbiased regardless of the heteroskedasticity structure. Also notice that all these estimators are asymptotically equivalent. Therefore, we can apply Theorem 4.2.7 (ii) in order to obtain—large sample confidence intervals for the unknown slope parameter β.

In finite sample considerations assuming the errors are normally distributed, one can obtain four $(1-\alpha)100\%$ confidence intervals for β, each of whose *target* coverage is $(1-\alpha)100\%$.

$$\hat{\beta}_n \mp t_{\alpha/2,\nu} \sqrt{\tilde{V}_{n,w}}, \tag{4.54}$$

$$\hat{\beta}_n \mp t_{\alpha/2,\nu} \sqrt{\tilde{V}_{n,mw}}, \tag{4.55}$$

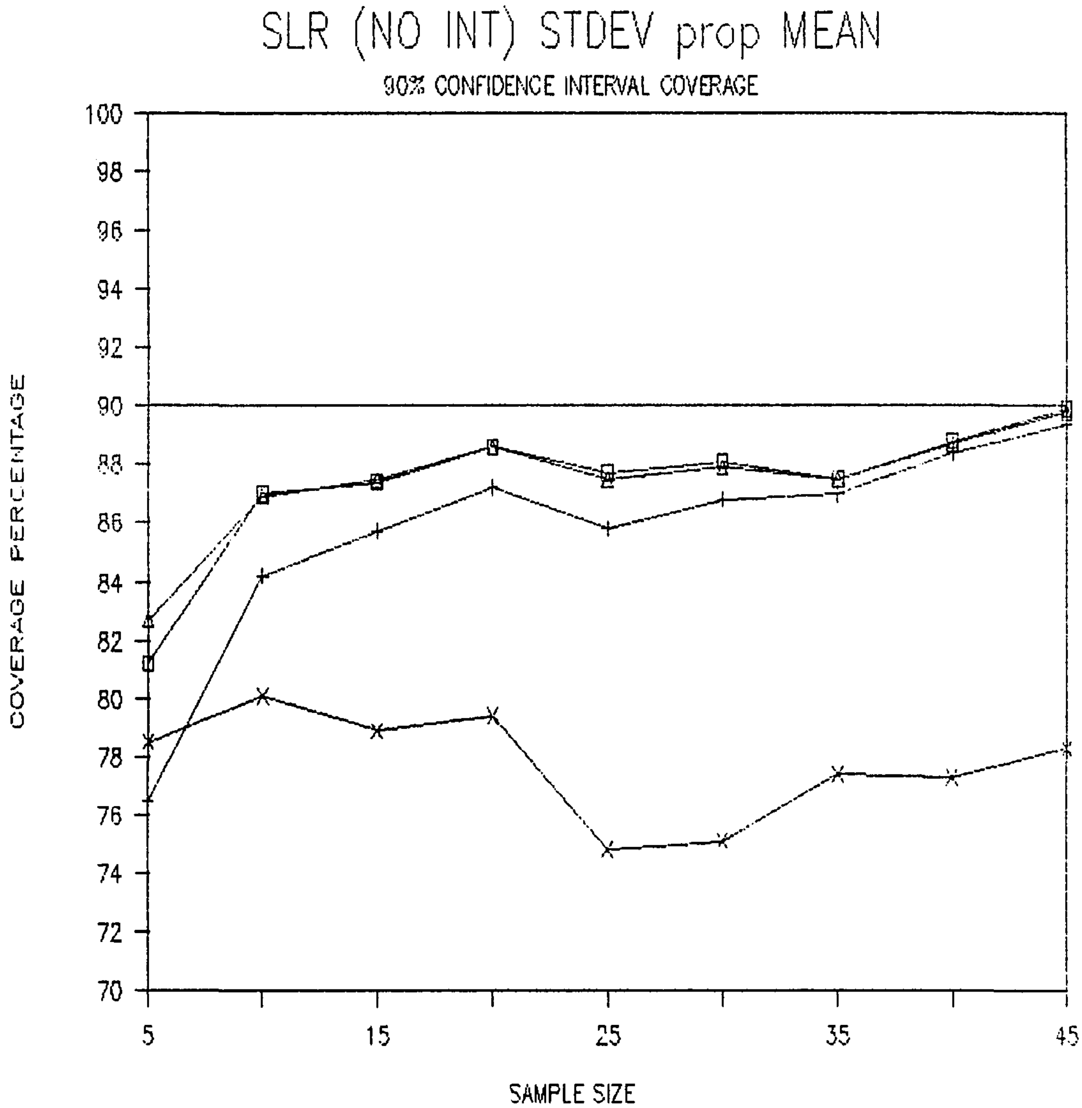
$$\hat{\beta}_n \mp t_{\alpha/2,\nu} \sqrt{\tilde{V}_{n,unb}}, \tag{4.56}$$

and

$$\hat{\beta}_n \mp t_{\alpha/2,\nu} \sqrt{\tilde{V}_{n,usu}}, \tag{4.57}$$

where $\hat{\beta}_n$ is the o.l.s estimator of β given by (4.44), $\tilde{V}_{n,w}$, $\tilde{V}_{n,mw}$, $\tilde{V}_{n,unb}$, and $\tilde{V}_{n,usu}$ are given by (4.51), (4.52), (4.53) and (4.48) respectively, and $t_{\alpha/2,\nu}$ is obtained from a Student's t—distribution with $\nu=(n-1)$ degrees of freedom where $\alpha/2$ is the probability that a t—random variable is greater than $t_{\alpha/2,\nu}$.

The performances of the confidence intervals above were studied via a Monte Carlo simulation. The design we adopted was closely related to the design used by Dorfman (1988). Data sets were generated according to the model $Y_i = X_i\beta + \epsilon_i$ $(i=1,2,\ldots,n)$, using the IMSL double precision normal random number generator DRNNOR. Throughout the simulation, the slope parameter was fixed at $\beta=1$. We considered 9 different sample sizes: 5, 10,15,...,45, and X_i's were chosen as $X_i = i$ $(i=1,2,\ldots,n)$. Two structures for heteroskedasticity were imposed: (i) $\sigma_i \propto$ mean and (ii) $\sigma_i \propto$ mean^2. For each n and $\{\sigma_i\}$ configuration the actual coverage probabilities were estimated according to the four different confidence intervals given by (4.54) through (4.57). We considered two target confidence coefficients: 90% and 95%. The results of the simulation are summarized in Figures 4.6 and 4.7. These plots show that none of the confidence intervals attains the specified confidence coefficient and all of them are liberal. We see that the confidence interval given by (4.57) and based on $\hat{V}_{n,usu}$, is very liberal in comparison to the other three confidence intervals. The confidence interval based on $\tilde{V}_{n,unb}$ performs slighlty better than the interval based on $\tilde{V}_{n,mw}$ and these two clearly perform better than the other two particularly in small samples.



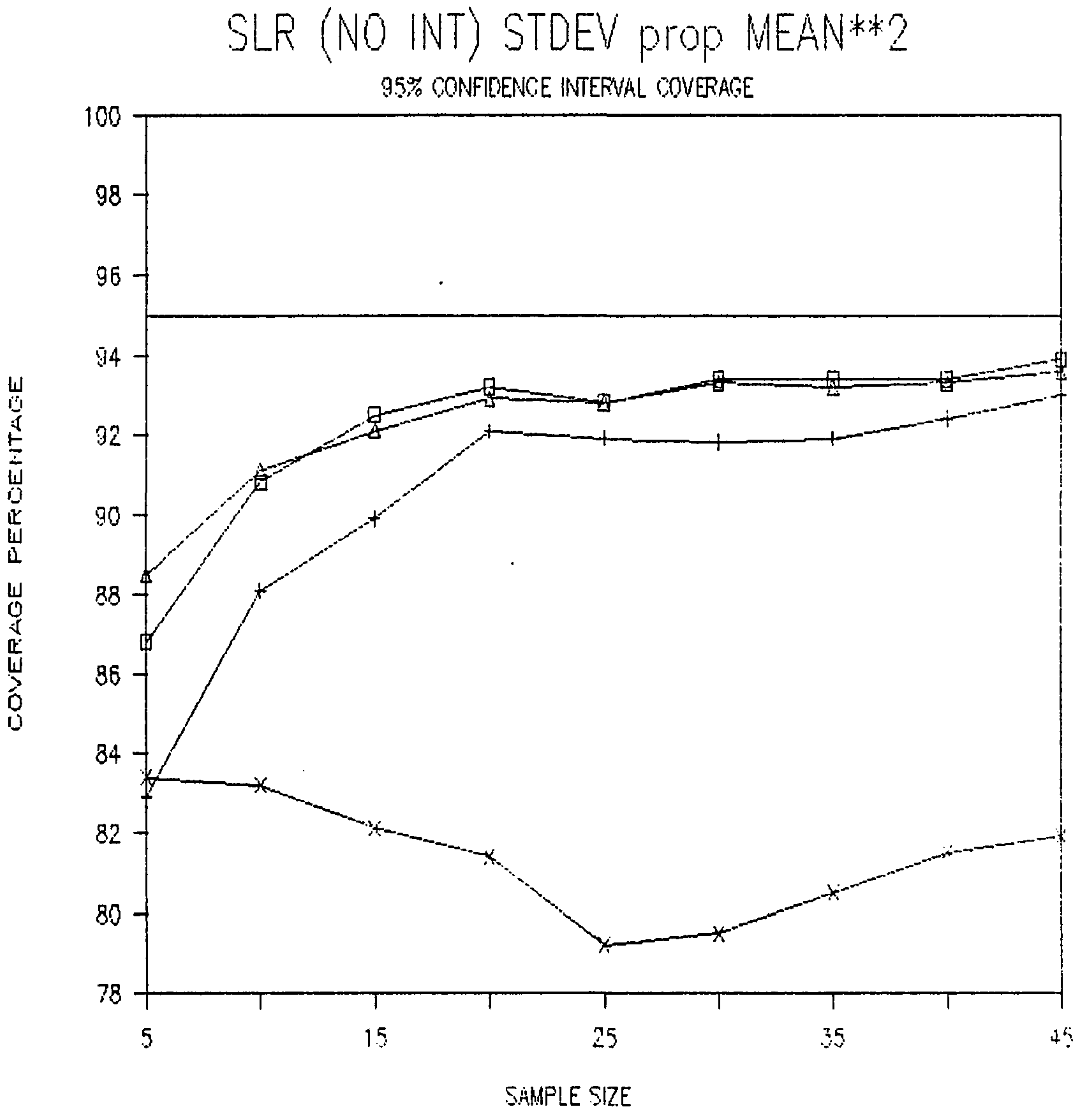
+ : corresponds to $\tilde{V}_{n,w}$

△ : corresponds to $\tilde{V}_{n,mw}$

□ : corresponds to $\tilde{V}_{n,unb}$

X : corresponds to $\tilde{V}_{n,usu}$

Figure 4.6. Simple linear regression with intercept, $\sigma_i \propto$ mean



+ : corresponds to $\tilde{V}_{n,w}$

Δ : corresponds to $\tilde{V}_{n,mw}$

$\square$: corresponds to $\tilde{V}_{n,unb}$

X : corresponds to $\tilde{V}_{n,usu}$

Figure 4.7. Simple linear regression with intercept,

$\sigma_i \propto \text{mean}^2$

4.5.2. Simple linear regression with intercept

Simple linear regression model with intercept can be written formally as,

$$Y_i = \alpha_o + \alpha_1 X_i + \epsilon_i, \tag{4.58}$$

where $E(\epsilon_i)=0$, and $E(\epsilon_i^2)=\sigma_i^2 < \infty$, $(i=1,2,\ldots,n)$. In matrix notation we can write

$$\underline{Y} = X\underline{\beta}_o + \underline{\epsilon},$$

where

$$\underline{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}_{n \times 1},$$

$$X = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix}_{n \times 1},$$

$$\underline{\beta}_o = \begin{bmatrix} \alpha_o \\ \alpha_1 \end{bmatrix},$$

and

$$\underline{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}_{n \times 1}.$$

Using the notation of Section 4.2, we see that

$$\bar{M}_n = \frac{1}{n}\begin{bmatrix} n & \sum_{i=1}^{n} X_i \\ \sum_{i=1}^{n} X_i & \sum_{i=1}^{n} X_i^2 \end{bmatrix},$$

$$\bar{V}_n = \frac{1}{n}\begin{bmatrix} \sum_{i=1}^{n} \sigma_i^2 & \sum_{i=1}^{n} \sigma_i^2 X_i \\ \sum_{i=1}^{n} \sigma_i^2 X_i & \sum_{i=1}^{n} \sigma_i^2 X_i^2 \end{bmatrix},$$

$$h_{ii} = (1, X_i)\,(X'X)^{-1}\,(1, X_i)',$$

$$= \frac{1}{\{n\sum_{j=1}^{n} X_j^2 - (\sum_{j=1}^{n} X_j)^2\}}\,(1, X_i)\begin{bmatrix} \sum_{j=1}^{n} X_j^2 & -\sum_{j=1}^{n} X_j \\ -\sum_{j=1}^{n} X_j & n \end{bmatrix}(1, X_i)'$$

$$= \frac{\sum_{j=1}^{n} X_j^2 - 2X_i \sum_{j=1}^{n} X_j + nX_i^2}{n\sum_{j=1}^{n} X_j^2 - (\sum_{j=1}^{n} X_j)^2}, \tag{4.59}$$

hence

$$(1 - h_{ii}) = \frac{(n-1)\sum_{j=1}^{n} X_j^2 - (\sum_{j=1}^{n} X_j)^2 + 2X_i \sum_{j=1}^{n} X_j + nX_i^2}{n\sum_{j=1}^{n} X_j^2 - (\sum_{j=1}^{n} X_j)^2},$$

and the ordinary least squares (o.l.s.) estimator of $\underline{\beta}_0 = \begin{bmatrix} \alpha_0 \\ \alpha_1 \end{bmatrix}$ is given by

$$\hat{\alpha}_1 = \frac{\sum_{j=1}^{n} X_j Y_j - (\sum_{j=1}^{n} X_j)(\sum_{j=1}^{n} Y_j)/n}{\sum_{j=1}^{n} X_j^2 - (\sum_{j=1}^{n} X_j)^2/n}, \tag{4.60}$$

$$\hat{\alpha}_O = \frac{1}{n}\left(\sum_{j=1}^{n} Y_j - \hat{\alpha}_1 \sum_{j=1}^{n} X_j\right). \tag{4.61}$$

Therefore the o.l.s. residuals are given by

$$\hat{\epsilon}_i = Y_i - \hat{\alpha}_O - \hat{\alpha}_1 X_i, \quad (i=1,2,\ldots,n). \tag{4.62}$$

Again, as we considered earlier, in many situations we are interested in testing hypotheses concerning the slope parameter α_1, i.e., we would like to test,

$$H_O : \alpha_1 = \alpha_{1O} \qquad \text{vs} \qquad H_1 : \alpha_1 \neq \alpha_{1O}$$

or in constructing confidence intervals for α_1. We can rewrite (4.60) as follows:

$$\hat{\alpha}_1 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})Y_i}{SS_{XX}}, \tag{4.63}$$

where

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n},$$

and

$$SS_{XX} = \sum_{j=1}^{n} X_j^2 - (\sum_{j=1}^{n} X_j)^2/n = \sum_{i=1}^{n}(X_i - \bar{X})^2.$$

From (4.62) we obtain

$$\text{var}(\hat{\alpha}_1) = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2 \sigma_i^2}{SS_{XX}^2}. \tag{4.64}$$

In order to obtain an estimator of $\mathrm{var}(\hat{\alpha}_1)$, White (1980a) suggested to replace σ_i^2 by the o.l.s. residual $\hat{\epsilon}_i$ given by (4.62). Thus we obtain an estimator $A_{n,w}$ of $\mathrm{var}(\hat{\alpha}_1)$ namely

$$A_{n,w} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2 \hat{\epsilon}_i^2}{SS_{XX}^2}. \tag{4.65}$$

In general $A_{n,w}$ is not an unbiased estimator of $\mathrm{var}(\hat{\alpha}_1)$. MacKinnon and White (1985) suggest the replacement of σ_i^2 in (4.63) instead by $\hat{\epsilon}_i^2/(1-h_{ii})$ in order to obtain an estimator of $\mathrm{var}(\hat{\alpha}_1)$ where h_{ii} is given by (4.59). Call this estimator $A_{n,mw}$; i.e.,

$$A_{n,mw} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2 \hat{\epsilon}_i^2/(1-h_{ii})}{SS_{XX}^2}. \tag{4.66}$$

The above estimator is also biased in general, but is unbiased in the homoskedastic error variance situation. Under the assumption of homoskedasticity, in the model (4.58) another estimator of $\mathrm{var}(\hat{\alpha}_1)$ is available (say $A_{n,usu}$). i.e.,

$$A_{n,usu} = \frac{\hat{\sigma}^2}{SS_{XX}}, \tag{4.67}$$

where

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n}\hat{\epsilon}_i^2}{(n-2)}.$$

Assuming the errors are normally distributed, one can obtain three $(1-\alpha)100\%$ confidence intervals for α_1 using the above three estimators of $\mathrm{var}(\hat{\alpha}_1)$ each of whose *target* coverage is $(1-\alpha)100\%$. Specifically,

$$\hat{\alpha}_1 \mp t_{\alpha/2,\nu} \sqrt{A_{n,w}}\,, \tag{4.68}$$

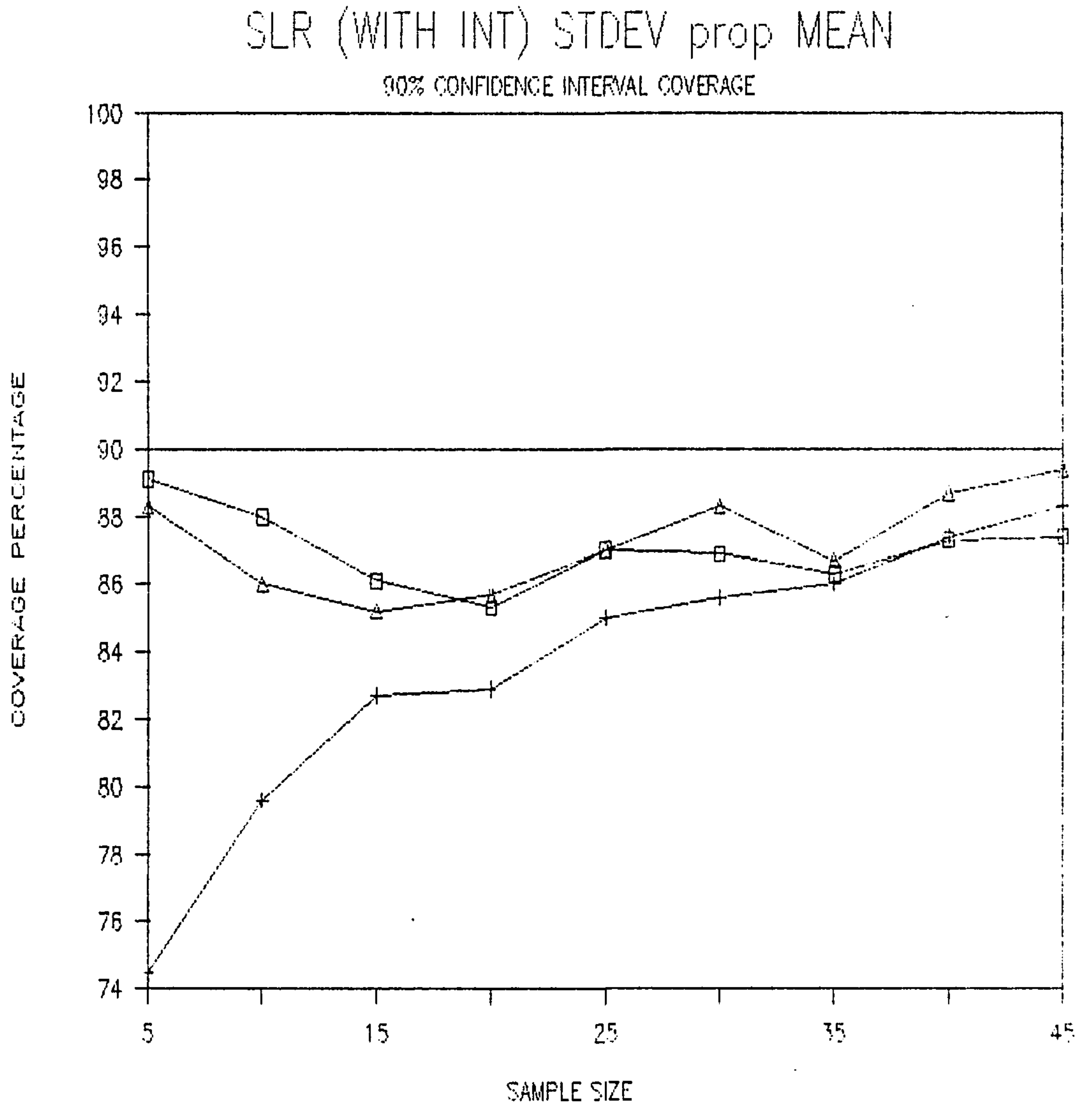
$$\hat{\alpha}_1 \mp t_{\alpha/2,\nu} \sqrt{A_{n,mw}}\,, \tag{4.69}$$

and

$$\hat{\alpha}_1 \mp t_{\alpha/2,\nu} \sqrt{A_{n,usu}}\,, \tag{4.70}$$

where $\hat{\alpha}_1$ is the o.l.s. estimator of α_1 given by (4.60), and $t_{\alpha/2,\nu}$ is obtained from a Student's t—distribution with $\nu = (n-2)$ degrees of freedom such that the probability that a t—random variable is greater than $t_{\alpha/2,\nu}$, is $\alpha/2$.

The performance of the confidence intervals above were studied via a simulation. A design similar to the one we discussed under simple linear regression without intercept was adopted. We set $\alpha_0 = 0$ and $\alpha_1 = 1$ throughout the simulation. The two target confidence coefficients were set at 90% and 95%. Two structures of heteroskedasticity were considered: (i) $\sigma_i \propto$ mean, and (ii) $\sigma_i \propto \text{mean}^2$. The results of the simulation are summarized in Figures 4.8 and 4.9. In both configurations of heteroskedasticity, we see that all three confidence intervals are liberal, but the one given by (4.69) shows a considerably better performance compared to the other two. For small samples when heteroskedasticity is present performance of the usual confidence interval given by (4.70) seems to be very poor.

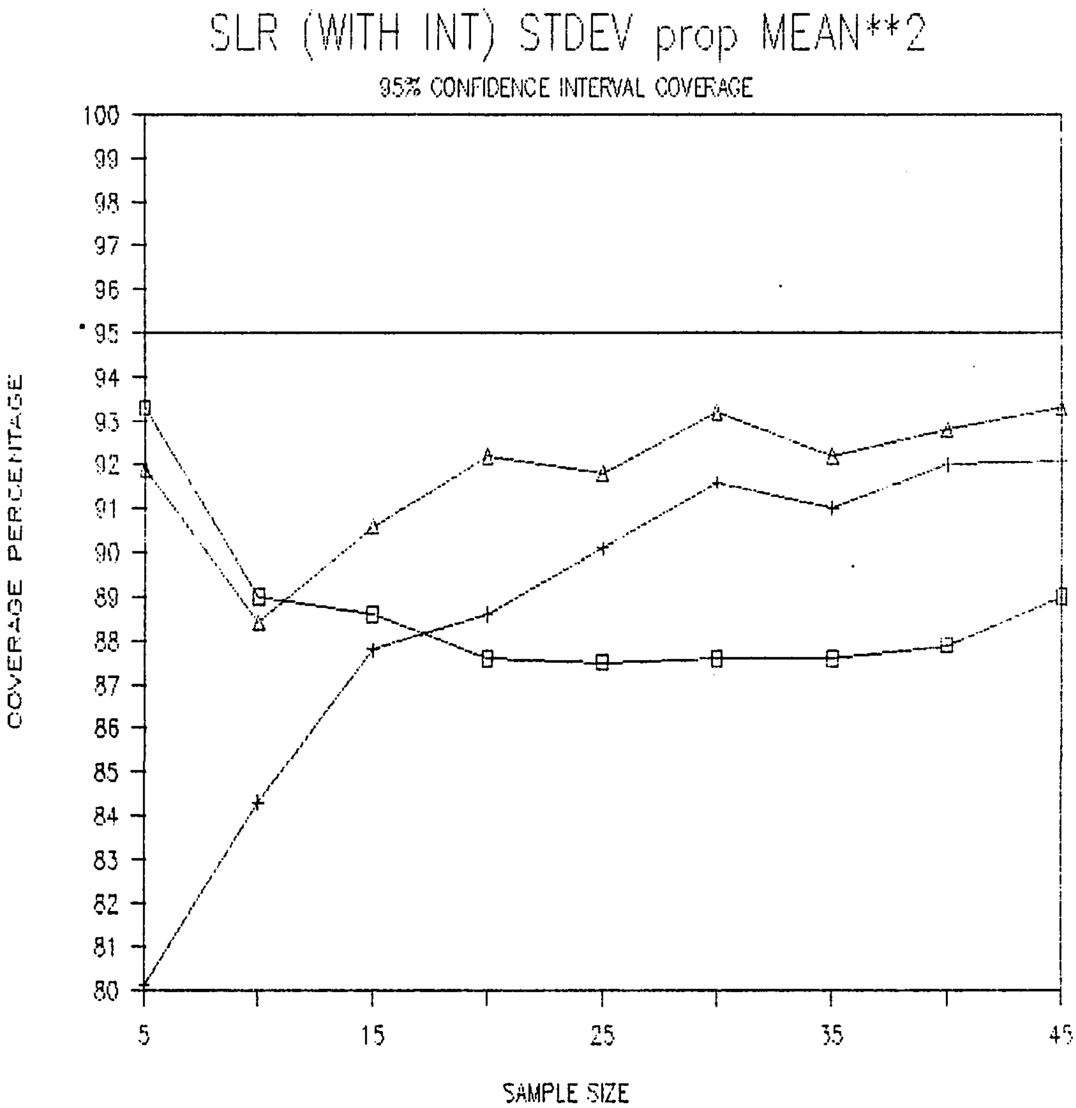


$+$: corresponds to $\tilde{V}_{n,w}$

$\triangle$: corresponds to $\tilde{V}_{n,mw}$

$\square$: corresponds to $\tilde{V}_{n,usu}$

Figure 4.8. Simple linear regression with intercept, $\sigma_i \propto$ mean



$+$: corresponds to $\tilde{V}_{n,w}$

$\triangle$: corresponds to $\tilde{V}_{n,mw}$

$\square$: corresponds to $\tilde{V}_{n,usu}$

Figure 4.9. Simple linear regression with intercept, $\sigma_i \propto \text{mean}^2$

4.5.3. Example

In this subsection we analyse the data originally presented by Williams (1959). The data set (see Table 1) consists of 11 observations on failing stress (modulus of rupture) f_θ of timber at an angle θ to the grain direction.

θ	f_θ
0	16,880
2.5	14,720
5	14,340
7.5	12,740
10	12,390
15	7,140
20	7,170
30	4,710
45	2,280
60	1,720
90	970

Table 1.

The relationship between θ and f_θ is rather complicated, but this may be reduced to a linear form by using Hankinson's formula. This relationship is as follows.

$$f_\theta = \frac{f_c f'_c}{f_c \sin^2\theta + f'_c \cos^2\theta}, \tag{4.71}$$

where f_c is the stress parallel to the grain direction and f'_c is the stress perpendicular to the grain direction. One can rearrange (4.71) can be written as

$$\frac{1}{f_\theta} = \frac{1}{f_c} + \sin^2\theta\left(\frac{1}{f'_c} - \frac{1}{f_c}\right). \tag{4.72}$$

If we let $Y = \frac{1}{f_\theta}$ and $X = \sin^2\theta$ then (4.72) takes the form

$$Y = \alpha_o + \alpha_1 X \ . \tag{4.73}$$

Therefore if we obtain estimates of α_o and α_1 then we can obtain the estimates of f_c and f'_c; namely $\hat{f}_c = \frac{1}{\hat{\alpha}_o}$ and $\hat{f}'_c = \frac{1}{(\hat{\alpha}_o + \hat{\alpha}_1)}$. In this analysis we consider $Y = \frac{10^6}{f_\theta}$ for computational simplicity. The corresponding transformed values Y and X are given in the following Table 2.

Y	X
59.24	0000.0
67.93	0.0019
69.74	0.0076
78.48	0.0170
80.71	0.0302
140.10	0.0670
139.50	0.1170
212.30	0.2500
438.60	0.5000
581.40	0.7500
1031.00	1.0000

Table 2.

Figure (4.10) shows the scatter plot of Y vs X. This plot indicates that the variability in Y increases as X increases, and clearly a heteroskedastic situation. Although it seems one can model the variance of Y proportional to some power of the mean we will not consider such approaches as it is not the scope of this dissertation. We will be looking at situations where such detailed modelling of variance function is not feasible. Ordinary least squares estimates of α_o and α_1 are found to be

$$\begin{bmatrix} \hat{\alpha}_o \\ \hat{\alpha}_1 \end{bmatrix} = \begin{bmatrix} 48.181 \\ 864.311 \end{bmatrix}.$$

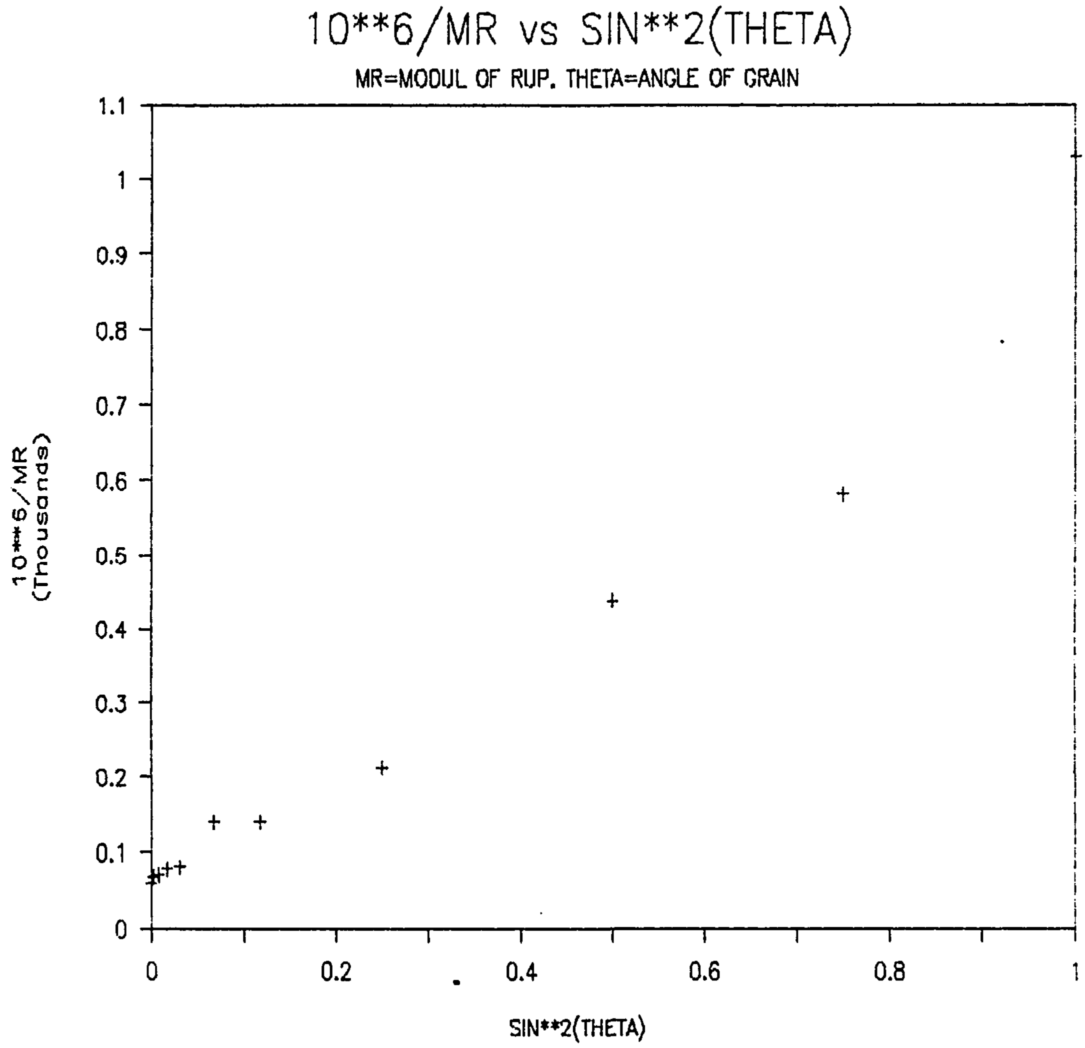


Figure 4.10. 10^6/modulus of rupture vs $\sin^2\theta$

Suppose we are interested in constructing a confidence interval for α_1. Here, as we discussed in the previous section we can obtain 3 estimates for the $\mathrm{var}(\hat{\alpha}_1)$. In the notation we used earlier, these estimates are found to be

$$A_{n,usu} = 3110.2896$$

$$A_{n,w} = 7788.0529$$

and

$$A_{n,mw} = 15521.8486$$

Therefore we could obtain the following 95% confidence intervals for α_1 using (4.68), (4.69), and (4.70).

usual : (738.162, 990.462)

White : (664.162, 1063.933)

M and W : (582.497, 1146.127) .

The interval based on MacKinnon and White's variance estimator seems to be the widest, whereas the interval based on the usual variance estimator seems to be the shortest. This reinforces our findings from the simulation study. There we found that when the variance of Y increase with X all three confidence intervals are liberal, but the interval based on MacKinnon and White's variance estimator is less liberal compared to the other two. Thus we would expect this interval to be wider, and from our simulation findings we are happy to use the interval given by (4.70).

In this chapter we looked at one—sample, two—sample, and simple linear regression problems which commonly arise in statistics, under the relaxed assumption of homoskedasticity. With regard to the one—sample and two—sample problems we found that if the heteroskedasticity is mild, i.e., $\max(\sigma^2)/\min(\sigma^2) < 1.5$ then the usual test procedures are quite robust. This was observed by looking at the loss of degrees of freedom. When the heteroskedasticity is quite severe we obtained equivalent degrees of freedom to approximate the distribution of the usual T—statistics. For the simple linear regression without intercept problem we constructed an exact unbiased estimator of the variance of the slope parameter and established the connection of this estimator to the results obtained by Cressie (1982) in connection to the one—sample problem. We also conducted a simulation study and found that the performance of this unbiased estimator is superior to the other existing estimators in constructing confidence intervals for the slope parameter. For the simple linear regression with intercept problem, we showed via simulation that MacKinnon and White's (1985) results shows superiority when compared to the other methods. Constructing exact unbiased estimators for the variance of parameter estimates of the simple linear regression with intercept problem is under investigation, as is the more general problem of linear regression.

5. CONCLUSIONS

In this dissertation we discussed a number of common statistics problems that we come across in statistics, namely the problem of combining unbiased estimators, the one—sample problem, the two—sample problem, and the linear regression problem. The primary goal of this dissertation was to relax the homoskedasticity assumption that is usually made in search of solutions to these problems.

An extensive literature review is given Chapter 1. In Chapter 2, robustness of the usual T—test and the effects of the parent population on the distributional properties of T was studied using Edgeworth expansions. Our studies confirmed earlier findings in the literature that were presented in Chapter 1. As the usual T—statistic is greatly influenced by the skewness of the parent population, we discussed a modification to the T—statistic using Cornish—Fisher expansions, suggested by Johnson (1978), and corrected the misprints in his article.

Chapter 3 was devoted to weighted estimation. Weighted estimation of an unknown parameter plays an important role when the observations are taken with different precision. We presented two theorems on weighted means when the chosen weights are random and two open problems of generalizing these theorems arose. The idea of M—estimation and weighted M—estimation was discussed in the same chapter and some results on forming asymptotically *safe* test statistics using weighted M—estimators were also given. Finite sample distributional properties of these test statistics need to be studied in the future. Application of the notion of equivalent degrees of freedom to the suggested test statistics remains an open problem.

The linear model under heteroskedastic errors was discussed in Chapter 4. The theme of this chapter was the application of White's (1980a) results to one—sample, two—sample and simple linear regression problems. We looked at the finite sample properties of the usual T—statistics and gave specific formulae for the equivalent degrees of freedom in approximating the finite sample distributions of these T—statistics by t—distributions.

6. REFERENCES

Abbott, J. H. and Rosenblatt, J. I. 1963. Two stage estimation with one observation in the first stage. Ann. Inst. Statist. Math. XIV: 229—235.

Amemiya, T. 1973. Regression analysis when the variance of the dependent variable is proportional to the square of its expectation. J. Am. Statist. Assoc. 68: 928—934.

Andrews, D. F., Bickel, P. J., Hampel, F. R., Huber, P. J., Rogers, W. H., and Tukey, J. W. 1972. Robust estimates of location: Survey and advances. Princeton University Press, Princeton, New Jersey.

Anh, V. V. 1988. Nonlinear least squares and maximum likelihood estimation of a heteroscedastic regression model. Stochastic Process. Appl. 29: 317—333.

Aucamp, D. C. 1986. A test for the difference of means. J. Statist. Comput. Simulation 24: 33—46.

Banerjee, S. 1960. Approximate confidence intervals for linear functions of means of k—populations when the population variances are not equal. Sankhyá 22: 357—358.

Banerjee, S. 1961. On confidence interval for two—mean problem based on seperate estimates of variances and tabulated values of t—table. Sankhyá 23(A): 359—378.

Bartlett, M. S. 1935. The effects of non—normality on the t distribution. Proc. of the Cambridge Philos. Soc. 31: 223—231.

Bell, W. W. 1968. Special functions for scientists and engineers. D. Van Nostrand Co. LTD, London.

Belsey, D. A., Kuh, E., and Welsch, R. E. 1980. Regression diagnostics. John Wiley and Sons, New York.

Bement, T. R. and Williams J. S. 1969. Variance of weighted regression estimators when sampling errors are independent and heteroscedastic. J. Am. Statist. Assoc. 64: 1369—1382.

Benjamini, Y. 1983. Is the t test really conservative when the parent distribution is long—tailed? J. Am. Statist. Assoc. 78: 645—654.

Bhattacharya, C. G. 1980. Estimation of a common mean and recovery of interblock information. Ann. Statist. 8: 205—211.

Bhattacharya, C. G. 1984. Two inequalities with an application. Ann. Inst. Statist. Math. 36, Part A, : 129—134.

Bhattacharya, R. N. and Ghosh, J. K. 1978. On the validity of the formal Edgeworth expansion. Ann. Statist. 6: 435—451.

Bickel, P. J. 1978. Using residuals robustly I: Tests for heteroscedasticity, nonlinearity. Ann. Statist. 6: 266—291.

Bickel, P. J. and Doksum, K. A. 1981. An analysis of transformations revisited. J. Am. Statist. Assoc. 76: 296—311.

Birch, J. B. and Binkley, D. A. 1983. Effects of non—normality and mild heteroscedasticity on estimators in regression. Commun. Statist. B—Simulation Comput. 12(3): 331—354.

Blachman, N. M. and Machol, R. E. 1987. Confidence intervals based on one or more observations. IEEE Transactions on Information Theory IT—33, No. 3: 373—382.

Bondesson, L. 1983. When is the t—statistic t—distributed. Sankhyá. 45(A): 338—345.

Boneau, C. A. 1960. The effects of violations of assumptions underlying the t—test. Psychological Bulletin 57: 49—64.

Bowman, K., Beauchamp, J. and Shenton, L. 1977. The distribution of the t—statistic under non—normality. Int. Statist. Rev. 45: 233—242.

Box, G. E. P. 1954. Some theorems on quadratic forms applied in the study of analysis of variance problems. I. Effect of inequality of variance in the one—way classification. Ann. Math. Statist. 25: 290—302.

Box, G.E. P. and Anderson, S. L. 1955. Permutation theory in the derivation of robust criteria and the study of departures from assumption. J. Roy. Statist. Soc. 17(B): 1—26.

Box, G. E. P. and Hill, W. J. 1974. Correcting inhomogeneity of variance with power transformation weighting. Technometrics 16: 385—389.

Box, G. E. P. and Watson, G. S. 1962. Robustness to non—normality of regression tests. Biometrika 49: 93—106.

Brown, M. B. 1982. Robustness against inequality of variances. Austral. J. Statist. 24: 283—295.

Brown, M. B. and Forsythe, A. B. 1974. The small sample behavior of some statistics which test the equality of several means. Technometrics 16: 129—132.

Callaert, H. and Veraverbeke, N. 1981. The order of the normal approximation for a Studentized U—statistic. Ann. Statist. 9: 194—200.

Carroll, R. J. 1980. A robust methods for testing transformations to achieve approximate normality. J. Roy. Statist. Soc. 42B: 71—78.

Carroll, R. J. and Cline, D. B. H. 1988. An asymptotic theory for weighted least—squares with weights estimated by replication. Biometrika 75: 35—43.

Carroll, R. J. and Ruppert, D. 1982. Robust estimation in heteroscedastic linear models. Ann. Statist. 10: 429—441.

Carter, E. M., Khatri, C. G., and Srivastava, M. S. 1979. The effect of inequality of variances on the t—test. Sankhyá 41B: 216—225.

Chaubey, Y. P. and Rao, P. S. R. S. 1976. Efficiencies of five estimators for the parameters of two linear models with unequal variances. Sankhyá 38B: 364—370.

Chew, V. 1970. Covariance matrix estimation in linear models. J. Am. Statist. Assoc. 65: 173—183.

Chung, K. L. 1946. The approximate distribution of Student's statistic. Ann. Math. Statist. 17: 447—465.

Chung, K. L. 1973. A course in probability theory. Academic Press, New York.

Cochran, W. G. 1964. Approximate significance levels of the Behrens—Fisher test. Biometrics 20: 191—195.

Cochran, W. G. and Carroll, S. P. 1953. A sampling investigation of the efficiency of weighting inversly as the estimated variance. Biometrics 9: 447—459.

Cochran, W. G. and Cox, G. M. 1957. Experimental designs. 2nd ed. John Wiley and Sons, New York.

Cohen, A. and Sackrowitz, H. B. 1984. Testing hypotheses about the common mean of normal distributions. Journal of Statistical Planning and Inference 9: 207—227.

Cook, R. D. and Weisberg, S. 1982. Residuals and Influence in regression. Chapman and Hall, New York.

Cornish, E. A. and Fisher, R. A. 1937. Moments and cumulants in the specification of distributions. Revue of the Int. Statist. Inst. 5: 307—320.

Cressie, N. A. C. 1980a. Relaxing assumptions in the one sample t—test. Aust. J. Statist. 22: 143—153.

Cressie, N. 1980b. M—estimation in the presence of unequal scale. Statistica Neerlandica 34: 19—32.

Cressie, N. A. C. 1982. Playing safe with misweighted means. J. Am. Statist. Assoc. 77: 754 —759.

Cressie, N. A. C. and Whitford, H. J. 1986. How to use the two sample t—test. Biometrical J. 28: 131—148.

Cressie, N.A. C., Sheffield, L. J. and Whitford, H. J. 1984. Use of the one sample t—test in the real world. J. Chronic Diseases 37: 107—114.

Dalal, S.R. 1978. Simultaneous confidence procedures for univariate and multivariate Behrens—Fisher type problems. Biometrika 65: 221—224.

Dalal, S. R., Tukey, J. W., and Cohen, M. L. 1984. Robust smoothly-heterogeneous variance regression. Pre print. Bell Communications Research.

Davidian, M. and Carroll, R. J. 1987. Variance function estimation. J. Am. Statist. Assoc. 82: 1079—1091.

Deaton, M. L., Reynolds, M. R., and Myers, R. J. 1983. Estimation and Hypothesis testing in regression in the presence of nonhomogeneous error variances. Comm. Statist. B—Simulation Comput. 12: 45—66.

Dorfman, A. 1988. A note on Miller's empirical weights for heteroskedastic linear regression. Comm. Statist. B—Simulation Comput. 17: 3521—3535.

Edgeworth, F. Y. 1898. On the representation of statistics by mathematical formulae. J. Roy. Statist. Soc. 61A: 670—700.

Efron, B. 1969. Student's t—test under symmetry conditions. J. Amer. Statist. Assoc. 63: 1278—1302.

Eicker, F. 1963. Asymptotic normality and consistency of the least squares estimators for families of linear regressions. Ann. Math. Statist. 34: 447—456.

Fuller, W. A. and Rao, J. N. K. 1978. Estimation for a linear regression model with unknown diagonal covariance matrix. Ann. Statist. 6: 1149—1158.

Games, P. A. and Howell, J. F. 1976. Pairwise multiple comparison procedures with unequal N's and/or variances: A Monte Carlo study. J. Educ. Statist. 1: 113—125.

Gans, D. J. 1981. Use of a preliminary test in comparing two sample means. Commun. Statist. B—Simulation. Comput. 10B: 163—174.

Gayen, A. K. 1949. The distribution of Student's' t in random samples of any size drawn from non—normal universes. Biometrika 36: 353—369.

Geary, R. C. 1936. The distribution of Student's ratio for non—normal samples. J. Roy. Statist. Soc. 2: 178—184.

Geary, R. C. 1966. A note on residual heterovariance and estimation efficiency in regression. Am. Statist. 20: 30—31.

Goldfeld, S. M. and Quandt, R. E. 1972. Non linear methods in econometrics. North Holland, Amsterdam.

Graybill, F. A. and Deal, R. B. 1959. Combining unbiased estimators. Biometrics 15: 543—550.

Groeneveld, R. and Meeden, G. 1977. The mode, median, and mean inequality. Am. Statist. 31: 120—121.

Gross, A. M. 1976. Confidence interval robustness with long—tailed symmetric distributions. J. Am. Statist. Assoc. 71: 409—417.

Hall, P. 1987. Edgeworth expansion for Student's t statistic under minimal moment conditions. Ann. Probab. 15: 920—931.

Hannan, E. J. 1971. Nonlinear time—series regression. J. Appl. Prob. 8: 767—780.

Hartley, H. O., Rao, J. N. K., and Kiefer, G. 1969. Variance estimation with one unit per stratum. J. Am. Statist. Assoc. 64: 841—851.

Havlicek, L. L. and Peterson, N. 1974. Robustness of the t—test: A guide for researches on effects of violations of assumptions. Psychological Reports 34: 1095—1114.

Hinkley, D. V. 1977. Jackknife in unbalanced situations. Technometrics 19: 285—292.

Hochberg, Y. 1976. A modification of the T—method of multiple comparisons for a one—way layout with unequal variances. J. Amer. Statist. Assoc. 71: 200—203.

Hotelling, H. 1961. The behavior of some standard statistical tests under non—standard conditions. Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability. 1. University of California Press, California, 319—359.

Huber, P. J. 1964. Robust estimation of a location parameter. Ann. Statist. 35: 73–101.

Huber, P. J. 1970. Studentizing robust estimates. In nonparametric techniques in statistical inference, ed. M. L. Puri. Cambridge University Press, Cambridge, U. K. 453–463.

Huber, P. J. 1981. Robust statistics. John Wiley and Sons, New York.

Jacquez, J. A. and Norusis, M. 1973. Sampling experiments on the estimation of parameters in heteroscedastic linear regression. Biometrics 29: 771–779.

Jacquez, J. A., Mather, F. J., and Crawford, C. R. 1968. Linear regression with non–constant, unknown error variances: Sampling experiments with least squares, weighted least squares and maximum likelihood estimators. Biometrics 24: 607–627.

James, G. S. 1951. Tests of linear hypotheses in univariate and multivariate analysis when the ratios of the population variances are unknown. Biometrika 38: 19–43.

Jennrich, R. I. 1969. Asymptotic properties of non–linear least squares estimators. Ann. Math. Stat. 40: 633–643.

Jobson, J. D. and Fuller, W. A. 1980. Least squares estimation when the covariance matrix and parameter vector are functionally related. J. Am. Statist. Assoc. 75: 176–181.

Johnson, N. J. 1978. Modified t test and confidence intervals for asymmetrical populations. J. Am. Statist. Assoc. 73: 536–544.

Kafadar, K. 1982. A biweight approach to the one sample problem. J. Am. Statist. Assoc. 77: 416–424.

Kantorovich, L. 1948. Functional analysis and applied mathematics. Usphehi Mat. Nauk, 3: 89–185. [English trans: Nat. Bur. Standards, Report No. 1509 (1952)]

Kendall, M. G. and Stuart, A. 1963. The advanced theory of statistics. Vol. 1. 2nd Edition. Griffin, London.

Khatri, C. G. and Shah, K. R. 1974. Estimation of location parameters from two linear models under normality. Comm. Statist. 3: 647–663.

Koenker, R. and Basset, G. Jr. 1982. Robust tests for heteroscedasticity based on regression quantiles. Econometrica 50: 43–46.

Kubokawa, T. 1987. Estimation of a common mean with symmetrical loss. J. Japan Statist. Soc. 17: 75–79.

Lauer, G. N. 1971. Powers of Cochran's test in Behrens—Fisher problems. Unpublished Ph.D. dissertation. Iowa State University, Ames, Iowa.

Machol, R. E. and Rosenblatt, J. 1966. Confidence interval based on single observation. Proc. IEEE 54: 1087—1088.

MacKinnon, J. G. and White, H. 1985. Some heteroskedasticity—consistent covariance matrix estimators with improved finite sample properties. J. Econometrics 29: 305—325.

Malinvaud, E. 1970. The consistency of nonlinear regressions. Ann. Math. Statist. 41: 956—969.

Meier, P. 1953. Variance of a weighted mean. Biometrics 9: 59—73.

Miller, R. G., Jr. 1986. Beyond Anova. Basics of applied statistics. John Wiley and Sons, New York.

Mosteller, F. and Tukey, J. W. 1977. Data analysis and regression: A second course in statistics. Addison—Wesley, Reading, Mass.

Nair, A. N. K. 1941. On the distribution of Student's "t" and the correlation coefficient in samples from non—normal populations. Sankhyá 5: 383—400.

Nair, K. A. 1980. Variance and distribution of the Graybill—Deal estimator of the common mean of two normal populations. Ann. Statist. 8: 212—216.

Nair, K. R. 1944. The recovery of inter—block information in incomplete block design. Sankhyá 6: 383—390.

Neyman, J. and Pearson, E. 1928. On the use and interpretation of certain test criteria for purposes of statistical inference, Part 1. Biometrika 20A: 175—240.

Norwood, T. E. and Hinkelmann, K. 1977. Estimating the common mean of several normal populations. Ann. Statist. 5: 1047—1050.

Ord, J. K. 1972. Families of frequency distributions. In Griffin's Statistical Monographs & Courses. Ed. Alan Stuart. Charles Griffin & Co. LTD, London.

Patel, K. R., Mudholkar, G. S., and Fernando, J. L. I. 1988. Student's t—approximations for three simple robust estimators. J. Am. Statist. Assoc. 83: 1203—1210.

Posten, H. O. 1979. The robustness of the one sample t—test over the Pearson system. J. Statist. Comput. Simulation 6: 133—149.

Prais, S. J. and Houthakker, H. S. 1955. The analyses of family budgets. Cambridge University Press, New York.

Rao, C. R. 1947. General method for analysis of incomplete block designs. J. Am. Statist. Assoc. 42: 541—561.

Rao, C. R. 1956. On the recovery of inter—block information in varietal trials. Sankhyá 17: 105—114.

Rao, C. R. 1970. Estimation of heteroscedastic variances in linear models. J. Am. Statist. Assoc. 65: 161—172.

Rao, C. R. 1973. Linear statistical inference and its applications. John Wiley and Sons Inc. New York.

Rao, J. N. K. 1973. On the estimation of heteroscedastic variances. Biometrics 29: 11—24.

Rao, J. N. K. and Subrahmanian, K. 1971. Combining independent estimators and estimation in linear regression with unequal variances. Biometrics 27: 971—990.

Rutemiller, H. C. and Bowers, D. A. 1968. Estimation in heteroscedastic linear model. J. Am. Statist. Assoc. 63: 552—557.

Satterthwaite, R. E. 1946. An approximate distribution of estimates of variance components. Biometrics 2: 110—114.

Scheffe, H. 1959. The analysis of variance. John Wiley and Sons, New York.

Scheffe, H. 1970. Practical solutions of the Behrens—Fisher problem. J. Am. Statist. Assoc. 65: 1501—1508.

Seshadri, V. 1963a. Constructing uniformly better estimators. J. Am. Statist. Assoc. 58: 172—178.

Seshadri, V. 1963b. Combining unbiased estimators. Biometrics 19: 163—169.

Shah, K. R. 1964. Use of inter—block information to obtain uniformly better estimators. Ann. Math. Statist. 85: 1064—1078.

Shao, J. 1988. Asymptotic theory in heteroscedastic nonlinear models. Technical Report, #88—2, Purdue University, Lafayette, Indiana.

Shao, J. and Wu, C. F. J. 1987. Heteroscedasticity—robustness of jackknife variance estimators in linear models. Ann. Statist. 15: 1563—1579.

Shinozaki, N. 1978. A note on estimating the common mean of K normal distributions and Stein problem. Commun. Statist.—Theor. Methods 7A: 1421—1432.

Smith, H. F. 1936. The problem of comparing results of two experiments with unequal error. J. Counc. Sci. Ind. Res. 9: 211—212.

"Sophister" 1928. Discussion of small samples drawn from an infinite skew population. Biometrika 20A: 389—423.

Spjotvoll, E. 1972. Joint confidence intervals for all linear functions of means in ANOVA with unknown variances. Biometrika 59:684—685.

Stein, C. 1966. An approach to the recovery of inter—block information in balanced incomplete block designs. *In Festschrift for J. Neyman*, F. N. David (Ed.). John Wiley and Sons, New York.

Student 1908. The probable error of a mean. Biometrics 6: 1—25.

Tamhane, A. C. 1977. Multiple comparisons in model I one—way ANOVA with unequal variances. Comm. Statist.—Theor. Methods 6A: 15—32.

Tamhane, A. C. 1979. A comparison of procedures for multiple comparisons of means with unequal variances. J. Am. Statist. Assoc. 74: 471—480.

Tan, W. Y. and Tabatabai, M. A. 1986. Some Monte Carlo studies on the comparison of several means under heteroscedasticity and robustness with respect to departures from normality. Biometrica J. 28: 801—814.

Tukey, J. W. 1964. Data analysis and behavioral sciences. Unpublished Manuscript. Princeton University, Princeton, New Jersey.

Tukey, J. W. and McLaughlin, D. H. 1963. Less vulnerable confidence and significance procedures for location based on a single sample: Trimming/ Winsorization I. Sankhyá 25A: 331—352.

Ury, H.K. and Wiggins, A. D. 1971. Large sample and other multiple comparisons among means. British J. Mathematical and Statistical Psycol. 24: 174—194.

Voinov, V. G. 1984. Variance and its unbiased estimator for the common mean of several normal populations. Sankhyá 46B: 291—300.

Wang, Y. Y. 1971. Probabilities of the Type I errors of the Welch tests for the Behrens—Fisher problem. J. Am. Statist. Assoc. 66: 605—608.

Wallace, D. L. 1958. Asymptotic approximations to distributions. Ann. Math. Statist. 29: 635—654.

Weber, N. C. 1986. The jackknife and heteroskedasticity. Consistent variance estimation for regression models. Economics Letters 20: 161—163.

Welch, B. L. 1937. The significance of the difference between two means when the population variances are unequal. Biometrika 29: 350—362.

Welch, B. L. 1951. On the comparison of several mean values: An alternative approach. Biometrika 38: 330—336.

White, H. 1980a. A heteroskedasticity—consistent covariance matrix estimator and a direct test for heteroskedasticity. Econometrica 48: 817—838.

White, H. 1980b. Nonlinear regression on cross—section data. Econometrica 48: 721—746.

White, H. and Domawitz, I. 1984. Nonlinear regression with dependent observations. Econometrica 52: 143—161.

Williams, E. J. 1959. Regression Analysis. John Wiley and Sons, New York.

Williams, J. S. 1967. The variance of weighted regression estimators. J. Am. Statist. Assoc. 62: 1290—1301.

Williams, J. S. 1975. Lower bounds on convergence rates of weighted least squares best linear unbiased estimators. A survey of statistical design and linear models. Ed. J. N. Srivastava, 555—570.

Wu, C. F. 1979. Asymptotic theory of nonlinear least squares estimator. Ann. Statist. 9: 501—513.

Wu, C. F. J. 1986. Jackknife, bootstrap and other resampling methods in regression analysis. Ann. Statist. 14: 1261—1295.

Yates, F. 1939. The recovery of intra—block information in varietal trial arranged in three dimensional latices. Ann. Eugenics 9: 136—156.

Yates, F. 1940. The recovery of inter—block information in balanced incomplete block designs. Ann. Eugenics 10: 317—325.

Yuen, K. K. 1974. The two sample trimmed t for unequal population variances. Biometrika 61: 165—170.

Yuen, K. K. and Murthy, V. K. 1974. Percentage points of the t—statistic when the parent is Student's t. Technometrics 16: 495—497.

6. ACKNOWLEDGEMENTS

I sincerely express my deepest gratitude and appreciation to Professor Noel A. C. Cressie who served as my advisor and major professor throughout my duration at Iowa State University and who suggested the topic and directed the research towards this dissertation. He also offered continous encouragement and assistance during this period.

I am thankful to many others, especially the members of my committee: Dr. Dean Isaacson, Dr. Richard Groeneveld, Dr. Stephen Vardeman, Dr. Robert Stephenson, and Dr. Wolfgang Kliemann. In addition, I am indebted to Dr. Clifford Bergman for serving as proxy committee member during Dr. Kliemann's absence.

Finally, I wish to thank my father, brothers, and sisters who encouraged me during my stay here at Iowa State.